



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

How rational and competitive is the market for mutual funds?

Leippold, Markus ; Rüegg, Roger

Abstract: To explore the rationality and competitiveness of the mutual fund industry, we analyze the alpha of active and index mutual funds from a global sample of more than 60,000 equity and fixed income funds and test the null hypothesis that alphas to investors are zero. We distinguish between institutional and retail investors since there are significant differences in management fees, economies of scale, and information asymmetries between these two groups. Using a new robust statistical test, we cannot reject our null hypothesis for the vast majority of investment categories. We find that the average active fund has less exposure to traditional risk factors, but higher sensitivity to alternative risk premia. Fund persistence and the impact of size and fees add further support to our conclusion that the mutual fund industry is highly competitive, except for US domestic funds. This set of funds is excessively overfunded compared to other fund categories.

DOI: <https://doi.org/10.1093/rof/rfz011>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-171148>

Journal Article

Accepted Version

Originally published at:

Leippold, Markus; Rüegg, Roger (2020). How rational and competitive is the market for mutual funds? *Review of Finance*, 24(3):579-613.

DOI: <https://doi.org/10.1093/rof/rfz011>

How rational and competitive is the market for mutual funds?*

Markus Leippold[†] Roger Rueegg[‡]

June 2, 2019

Abstract

To explore the rationality and competitiveness of the mutual fund industry, we analyze the alpha of active and index mutual funds from a global sample of more than 60,000 equity and fixed income funds and test the null hypothesis that alphas to investors are zero. We distinguish between institutional and retail investors since there are significant differences in management fees, economies of scale, and information asymmetries between these two groups. Using a new robust statistical test, we cannot reject our null hypothesis for the vast majority of investment categories. We find that the average active fund has less exposure to traditional risk factors, but higher sensitivity to alternative risk premia. Fund persistence and the impact of size and fees add further support to our conclusion that the mutual fund industry is highly competitive, except for US domestic funds. This set of funds is excessively overfunded compared to other fund categories.

JEL classification: C12, G10, G11, G20, G23

Keywords: Active investing, index investing, mutual funds, robust alpha test

*We are indebted to Jules van Binsbergen (the editor), Fabian Ackermann, Andreas Blöchliger, Vlad Dubikovskyy, Michel Habib, Thorsten Hens, Emmanuel Jurczenko, Andreas Kappler, Robert Kosowski, Alexandre Lauwers, Harald Lohre, Per Ostberg, Julien Penasse, James Sefton, Michael Wolf, and an anonymous referee for helpful comments. Furthermore, we thank the participants at the Frontiers of Factor Investing Conference 2018 at the University of Leicester, and the seminar participants at the University of Zurich, Zurich Cantonalbank, Unigestion Geneva, at the 2018 UBS Quant Conference in London, the 2018 Financial Data Science & Econometrics conference at the University of Loughborough, and the Paris December 2018 Finance Meeting EUROFIDAI - ESSEC. The content of this paper reflects the personal view of the authors. In particular, it does not necessarily represent the opinion of Zürcher Kantonalbank. This paper was previously circulated under the titles “Is active investing a zero-sum game after costs?” and “Fifty Shades of Active and Index Alpha.”

[†]Department of Banking and Finance, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland, email: markus.leippold@bf.uzh.ch.

[‡]Department of Banking and Finance, University of Zurich and Zurich Cantonal Bank, Hardstrasse 201, 8005 Zurich, Switzerland, email: roger.rueegg@zkb.ch.

1. Introduction

The emergence of index investing has led to a seemingly endless debate about the merits of active portfolio management. According to the logic of [Sharpe \(1991\)](#)’s active management arithmetic, active investing is doomed in aggregate, as [French \(2008\)](#) puts it. However, to escape this seemingly irrefutable conclusion, [Sharpe \(1991\)](#) leaves a back door open by pointing out three potential flaws in his theory. First, passive managers might not be genuinely passive. Second, there might be substantial differences among active managers. Third, the summary statistics of active managers might not truly represent the performance of the actively managed dollar.¹

In our analysis, we shed light on the potential pitfalls of Sharpe’s arithmetic using a rich dataset of 61,269 equity and fixed income funds that held USD 17.8 trillion assets under management by the end of 2016. The breadth of our data allows us to account for the heterogeneity of active asset managers by differentiating between institutional and retail funds, equity and fixed income funds, geographical regions, and investment categories. To explore managerial skill, investor rationality, and competitiveness of the market for mutual funds, we analyze the value-weighted performance before and after fees, benchmarking against investable indexes under an appropriate statistical framework.

Many research papers, investors, and advisors place themselves in either the active or passive camp. The staunch defenders of active investing argue along the lines of [Berk and Green \(2004\)](#), who show that rational markets do not contradict the existence of skilled fund managers who consistently beat the market. They build their argument on a fundamental

¹Some recent literature started to challenge Sharpe’s arithmetic. For instance, [Berk and van Binsbergen \(2015\)](#) correctly highlight that, even in Sharpe’s arithmetic, active mutual funds can outperform the market because they only represent a subset of active investors. [Pedersen \(2018\)](#) points to Sharpe’s implicit assumption that the market portfolio never changes, which does not hold in the real world, forcing passive investors to trade regularly. Recently, [van Binsbergen and Opp \(2019\)](#) criticizes that Sharpe’s arithmetic remains absolutely silent about the price discovery mechanism of active management. Moreover, Sharpe’s arithmetic also fails because even passive investors have to trade at least twice: once to get in and once to get out. Such trading patterns may be exploited by active investors. We thank Jules van Binsbergen for pointing this out.

principle of economics: agents earn economic rents if, and only if, they have a competitive advantage. Hence, active investing is a zero-sum game after fees. Recently, [Berk and van Binsbergen \(2015\)](#) have provided empirical support for the claim that mutual fund managers do have skills.

In contrast, the proponents of passive investing argue along the lines of [Fama and French \(2010\)](#) in that the high fees of active management turn it into a negative-sum game after costs. A negative net performance provides us with some indication of investor rationality. A statistically negative net performance implies that investors systematically commit too much money to a fund. Indeed, [French \(2008\)](#) and [Fama and French \(2010\)](#), among many others, provide ample evidence that actively managed US equity mutual funds underperform their multi-factor benchmark after fees. In their view, active investing is at most a zero-sum game before fees, but not after fees. Consequently, over recent years, we have witnessed a massive inflow of funds into index investing.

These observations naturally drive us to question the value of active management. Our main focus in answering this question will be on the performance comparison of active and index funds. In this sense, we leave aside the price discovery aspect of active management. Recently, [van Binsbergen and Opp \(2019\)](#) provide a framework that allows capturing the real economic value that active managers provide by reducing alphas in the market. Hence, index funds profit from the work of active managers, as they increase market efficiency and the value of the market portfolio. This price discovery mechanism provides yet another argument against Sharpe's arithmetic. Since we do not consider this potential free-rider problem by passive investors, our analysis can be regarded as a conservative estimate of the rationality and competitiveness of the mutual fund industry.

Before we can start with our performance analysis, we need to decide on how to measure performance and how to define an appropriate benchmark. As [Berk and van Binsbergen \(2017\)](#) convincingly argue, there is no unique way to measure performance: it depends on

the research question. If we want to assess the rationality of fund investors and the degree of competition in different markets, then the appropriate measure is the fund’s net alpha. If, however, we want to measure the skill of an active manager, then we need to calculate the value added, i.e., the fund’s assets under management times the gross alpha.

Typically, researchers use a well established multi-factor model to proxy for the alternative opportunity set available to investors. However, multi-factor models include long–short portfolios with often very high turnover, generating considerable transaction costs.² Furthermore, also as argued by [Berk and van Binsbergen \(2017\)](#), we might have a situation in which we measure the performance of a fund at a time when the fund manager would not have known about some factors, as they were identified only much later.

Nevertheless, [Fama and French \(2010\)](#) argue that benchmarking against multi-factor models leads to the same conclusions as benchmarking against index funds, because the value-weighted portfolio of index funds exhibits close to zero alphas. Although we agree with their arguments for the US equity market, many index funds can also exhibit negative alphas relative to a multi-factor benchmark, depending on the asset class and the market.³ Also, there is no general agreement on the factors that should be included in a benchmark, leading to a severe selection bias. Hence, for performance analysis, we favor suitably defined investable benchmarks, while we use multi-factor models to understand the potentially different risk and style exposures of active and index funds.

The choice of the benchmark is just as critical as the measurement of the active performance. As highlighted in [Berk and van Binsbergen \(2015\)](#) and [Pedersen \(2018\)](#), also passive managers have to trade, when they either enter or leave the market, or when changes in the

²For a recent approach to estimate mutual fund skill based on multi-factor benchmarks and Bayesian techniques, see [Andrikogiannopoulou and Papakonstantinou \(2016\)](#).

³In an earlier version of the paper, we also analyzed single funds and benchmarked them against different multi-factor models and an investable benchmark. We found that retail funds perform worse than institutional funds and that, surprisingly, also index funds may deliver negative alpha and perform even worse than active funds if benchmarked against multi-factor models. However, an index fund may also exhibit skill, in line with the results previously found in [Crane and Crotty \(2018\)](#).

replicating index occur.⁴ Hence, the first back door of Sharpe’s arithmetic becomes more important, since passive managers act not truly passive. Thus, to build the benchmark for active managers within an investment category, we value-weight all index investors within the same Morningstar investment category and thereby obtain the average opinion of all passive investor about the replicating index and the implementation methodology.⁵ Then, to analyze the rationality and competitiveness of the mutual fund industry, we focus on the value-weighted alpha of active funds within the Morningstar investment categories benchmarked against the value-weighted index funds within the same investment category.

In a preliminary analysis of our fund data, we find significant evidence for serial and cross-sectional dependence in our mutual fund data. Therefore, we use a statistical framework with two key elements. First, we develop a robust statistical test for the mutual funds’ alpha, which takes into account serial dependence. Second, we use these test statistics as input for the multiple hypothesis testing methods of [Romano and Wolf \(2016\)](#), which is robust to the presence of cross-sectional dependence.⁶ In a simulation study, we find that the standard inference techniques are liberal in rejecting the null hypothesis, while we obtain accurate empirical rejection probabilities for our block resampling based alpha test. Other papers that conduct bootstrapped inference, such as [Kosowski et al. \(2006\)](#) and [Fama and French \(2010\)](#), sample one-period returns and, therefore, lose any information on the potential serial dependence; nor do they adjust for multiple hypotheses.⁷

Using our robust alpha test as input for the method of [Romano and Wolf \(2016\)](#) to adjust for multiple hypotheses, we find that there are significant negative alphas after cost for the

⁴For instance, when MSCI Inc. decided to consider China A Shares in 2018 partially, all index investors that replicated their emerging market index started to buy the new inclusions.

⁵If we would take the best index fund or concentrate on a specific provider, we would include a selection bias in our comparison. Using a value-weighted approach, we obtain a representative benchmark for a given investment category.

⁶By doing so, we avoid the pitfalls of the multiple hypothesis approach of [Barras, Scalliet, and Wermers \(2010\)](#) based on the false discovery rate (FDR), which may produce significantly biased results in the presence of cross-sectional dependence as recently shown by [Andrikogiannopoulou and Papakonstantinou \(2018\)](#).

⁷The Matlab code of the robust statistical framework for the alpha is available from the authors on request.

institutional “US Equity Large Cap Blend” and retail “Canada Fixed Income” categories. This finding corroborates the conclusion of [Fama and French \(2010\)](#) for these specific investment categories. However, for all of the other categories, our results support [Berk and Green \(2004\)](#) and [Berk and van Binsbergen \(2015\)](#) in that we cannot reject the hypothesis that it is a zero-sum game after costs.⁸ Furthermore, before costs, none of the active fund categories generate significantly negative alphas. Some of them even show significantly positive alphas before costs.

Analyzing the drivers of the difference in the performance of active and index funds, we find that the equity and fixed income active managers have less exposure to traditional risk factors such as market and duration risk. Instead, active equity funds have a small cap and growth stock bias and active fixed income funds load on credit risk. Surprisingly, when the market is affected by unexpected volatility shocks, active management tends to underperform the average index investor. In periods of calm markets and when the implied volatility decreases, active managers tend to outperform. We explain this finding as being due to active managers who prefer to sell insurance and generate exposures to risk premia that perform well in good times but may cause substantial losses in bad times. The significant higher exposure to small cap companies for equity and credit risk for fixed income managers supports this hypothesis.⁹

Our data also allows us to shed light on the difference between retail and institutional funds and the role of fees. While before fees, retail and institutional active funds deliver a similar and positive alpha, the differences become large when we take fees into account. Particularly for fixed income funds, the aggregated alpha moves into negative territory. Moreover, we

⁸These findings resonate well with [Pastor, Stambaugh, and Taylor \(2015\)](#), who argue that in markets in which the mutual fund industry is big, such as the “US Equity Large Cap Blend,” active alphas tend to be negative, and the equal-weighted alpha within investment categories exceeds the value-weighted alpha.

⁹[Agarwal and Naik \(2004\)](#) find similar return patterns for hedge funds. Thus, mutual funds try to profit from the same opportunities as hedge funds but have of course a narrower set of investment opportunities, due to regulatory restrictions. However, our results seem to contradict some of the previous findings, such as those of [Moskowitz \(2000\)](#) and [Kosowski \(2011\)](#), among others. They find that actively managed mutual funds tend to perform better than their passive benchmarks in bad times. However, these papers do not cover the recent financial crisis.

find that small retail funds with high fees have a significantly negative alpha. Hence, we suspect that this observation is a primary driver of [Fama and French \(2010\)](#)’s conclusion, since these funds obtain a substantial weight in their performance analysis. Our results also provide direct evidence for [Garleanu and Pedersen \(2018\)](#), who argue that more sophisticated investors outperform small investors because of the higher economies of scale in searching for skilled active managers.¹⁰ Moreover, our results endorse the hypothesis of [Gennaioli, Shleifer, and Vishny \(2015\)](#), who claim that trust is an essential component of the high fees in asset management, and who argue that active retail managers profit from pandering to trusting investors by buying hot assets, which explains the tendency for active retail mutual funds to have positive exposure to growth stocks.

We further demonstrate, along the lines of [Carhart \(1997\)](#), that the average active retail investor can significantly improve their performance over the period ranging from 1993 to 2016, provided the worst-performing active mutual funds of the past year are neglected. However, when the investor concentrates only on the top performing funds, the overall performance cannot be significantly improved. Also, we explore the role of fund size and fees on fund performance. Sorting active fund portfolios according to their performance persistence, fees, and size, we find that winner portfolios with low-fee and small funds tend to outperform, but their alpha does not survive our test statistics. However, for both equity and fixed income retail funds, we find that a fund investor is well advised to avoid high-fee and small losers, as they generate significantly negative alphas.

We organize the remainder of this paper as follows. Section [2](#) discusses the data and performs a preliminary analysis, which motivates the design of our empirical tests. Section [3](#) presents our robust alpha test and the multiple hypothesis framework. In Section [4](#), we provide a comparison of the value-weighted performance of active and index mutual funds portfolios

¹⁰Also, our empirical analysis supports the findings of [Gerakos, Linnainmaa, and Morse \(2016\)](#), who show that institutional investors outperform their strategy benchmarks after fees. Their analysis is based on self-reported but GIPS (Global Investment Performance Standard) compliant data, which still may inherit some biases, while we build our analysis on publicly available performance data.

across investment categories and asset classes. We analyze the drivers of the performance difference of active and index funds, and we explore the role of performance persistence, fund size, and fees. Section 5 concludes.

2. Preliminary Analysis

We first describe our data. Then, we analyze the potential time and cross-sectional dependencies in mutual fund returns to guide the formulation of appropriate test statistics for our hypotheses.

2.1 Data

Our mutual fund sample is drawn from the Morningstar database and ranges from December 1991 to December 2016. We include a total of 61,269 funds from different asset classes.¹¹ Table 1 shows the summary statistics of cross-sectional monthly attributes across asset classes. In our data, we have 14,969 institutional and 46,300 retail funds, while we have 56,136 active and 5,133 index funds. In general, there are fewer index funds, but they show higher average total net assets (TNA) and net returns, and also lower fees and about the same average years in the database. As expected, the institutional funds charge lower fees than their retail counterparts.

[Table 1 about here.]

As of December 2016, the total net assets of equity retail funds amounted to USD 9 trillion, those of fixed income retail funds to USD 3.7 trillion, and those of equity institutional funds

¹¹In comparison, [Pastor, Stambaugh, and Taylor \(2015\)](#) explore 3,126 actively managed US equity-only mutual funds while [Berk and van Binsbergen \(2015\)](#) use 5,974 actively managed funds. [Ferreira et al. \(2013\)](#) analyze the performance of 16,316 open-end actively managed equity funds in 27 countries from 1997 to 2007. [Banegas et al. \(2013\)](#) focus on 4,200 European equity mutual funds. Hence, we add to the existing literature by providing evidence based on our new dataset. Furthermore, to the best of our knowledge, we are the first to apply a robust multiple hypothesis framework to active and index mutual funds in an international context and using investable benchmarks.

and fixed income institutional fund to USD 3.1 trillion and USD 2 trillion, respectively. Since institutional investors often invest their money through mandates, there are fewer institutional funds than retail funds. The assets under management for index funds have been steadily increasing over our sample period. By the end of 2016, we find the highest concentration of index funds for equity funds, with 28% for retail and 32% for institutional funds. Looking at the fixed income funds, we find 18% of the retail and 13% of the institutional funds were index funds. For a more detailed description of the data and the data cleaning procedures, we refer to Appendix [A](#).

For the performance analysis of fund portfolios, we construct investable benchmarks based on Morningstar’s investment categories.¹² We rely on these categories as they are well established in the industry, and their definition perfectly serves our intention to benchmark active funds. We acknowledge that there are many routes to take for benchmarking fund portfolios. In practice, when investors or active managers focus on a specific investment category, they do not compare themselves with the multi-factor models in general, rather, they compare themselves with other funds within the same category. As Morningstar states on its website, “the classifications were introduced in 1996 to help investors make meaningful comparisons between mutual funds.” While the investment objective stated in a fund’s prospectus does not always reflect how the fund actually invests, Morningstar places funds in a given category based on their portfolio statistics and securities holdings.

To construct an investable benchmark, we value-weight all index funds within a given investment category. By value-weighting the index funds in each category, we obtain our investable benchmarks which we use for calculating the alphas of active funds portfolios. In contrast to existing literature that focused on the Vanguard’s index funds only we include all

¹²See Table [A.1](#) in the Appendix for a list and summary statistics of the regarded investment categories. For the investable one-factor model, we require at least 12 monthly returns and for the multi-factor models at least 36 monthly returns.

index providers that operate within a specific investment category.¹³ We thereby reduce the selection bias that arises when we concentrate on one index provider only. Since we value-weight also the index funds, market leaders such as Vanguard still obtain a high weight in the analysis.

Since we require connected time-series for our multiple hypothesis adjustment, we focus on the periods from 1993 to 2016 and 2000 to 2016, which allows us to include more investment categories for the more recent time periods. Given that index mutual funds only emerged recently, we observe for the period starting in 1993 at least one index fund for four institutional and 17 retail categories. For the more recent period starting in 2000, we obtain 30 investment categories for the retail segment and 12 investment categories for the institutional segment. Hence, we end up with 63 categories.

2.2 Dependency Analysis

It is well known that statistical inference for econometric models is severely complicated by the existence of serial and cross-sectional dependencies. [Fama and French \(2010\)](#) find that cross-sectional dependence can materially change the inference and, therefore, propose an appropriate adjustment for their single fund analysis. At the same time, they correctly point to a potential caveat in their resampling approach. Because they perform a random sampling of months, they lose any effects of autocorrelation. Similarly, neither does [Barras, Scalliet, and Wermers \(2010\)](#) take into account serial dependence for their single-fund analysis, claiming that they find such an effect only for a few mutual funds. Moreover, they argue that the cross-sectional dependencies in their sample are sufficiently low to allow consistent estimators. Given that we do not analyze the returns of single funds as in [Barras, Scalliet, and Wermers \(2010\)](#) and [Fama and French \(2010\)](#) but of mutual fund portfolios, we must take a closer look at our data.

¹³As an example, [Fama and French \(2010\)](#) consider Vanguard’s index funds holding US stocks only. [Berk and van Binsbergen \(2015\)](#) include Vanguard’s index funds that hold not only US but also international stocks.

We first test for serial dependence, applying both the classical Ljung–Box (LJ) test and, as a robustness check, the distribution-free test of [Genest and Rémillard \(2004\)](#).¹⁴ For both tests we must fix the number of lags L , for which we use the automatic block-length selection for the dependent bootstrap of [Politis and White \(2004\)](#) and the correction of [Patton, Politis, and White \(2009\)](#). We find that most mutual funds show an optimal block size of two or three. Therefore, we set the lag L to three for the two tests. Testing for serial dependence, we find that on a 5% level the rejection rates range between 19% for retail funds and 44% for institutional funds. Furthermore, when we go into more detail, we find that the single mutual funds with the longest available time-series show a higher percentage of rejections. For example, the 2% oldest equity and fixed income single mutual funds, benchmarked against the investable model, exhibit statistically significant serial dependence in 40% and 63% of the cases. This evidence clearly justifies the need to control for dependence over time when we analyze the alpha of portfolios of mutual funds.

We next test for cross-sectional dependence, which might occur if mutual funds “herd” in their holdings, as is shown by [Wermers \(1999\)](#). To detect cross-sectional dependence in our data, we apply the test of [Pesaran \(2004\)](#). To compute the test statistic, we concentrate on funds that have more than one time period in common. Our empirical tests indicate that we can reject the null hypothesis of no cross-sectional dependence at the one percent significance level for all portfolio fund categories. While we find an average pairwise correlation of 14% (11%) for institutional (retail) funds, they vary across the different investment categories from -0.39% to as much as 66% . Since already low levels of cross-sectional and serial dependencies may distort statistical inference, the above empirical evidence dictates that statistical tests must take into account both serial and cross-sectional dependence.

¹⁴Much criticism has been leveled at the possible low power of the LJ test. The LJ test is based on autocorrelations and, hence, it is not a real test of independence. The test developed by [Genest and Rémillard \(2004\)](#) uses ranks and, therefore, is distribution-free and does not depend on the underlying distribution of the observations.

3. Robust Alpha Test and Multiple Hypotheses

To control for serial dependence, we propose a robust alpha test based on a studentized block bootstrap, which improves the accuracy of an inference for dependent time-series data compared to other methods. To compute the bootstrapped t -statistics and p -values we closely follow [Ledoit and Wolf \(2008, 2011\)](#), who study the related problem of testing whether two Sharpe ratios or two variances are equal. We outline the mathematical details of the bootstrapped standard error of the estimated alpha in [Appendix B](#). Once we have calculated the bootstrapped t -statistics and p -values in [Equations \(B.12\) and \(B.13\)](#), we can use them as input for multiple hypothesis testing.

While we control for serial dependence for the single-hypothesis alpha test, we control for cross-sectional dependence for the multiple-hypothesis method in which we control the family-wise error rate (FWER).¹⁵ When we control for the FWER, we apply the state of the art multiple hypothesis framework of [Romano and Wolf \(2016\)](#), which provides an efficient way to calculate the adjusted p -values. Since for the fund portfolios we have no missing values or disconnected time-series, as is the case for single funds, we can jointly sample blocks of fund and benchmark returns, thereby taking fully into account cross-sectional dependence. Since we sample the test statistics with our robust alpha test, we fully take into account the serial dependence structure as well.

To explore the accuracy of our test, we present the results of a simulation exercise in [Appendix C](#). We find that our robust alpha test is still liberal but more accurate since it also corrects for the serial dependence observed in the data. The standard inference tests are too liberal in rejecting the null hypothesis. Thus, when we apply the standard tests or sample only one return each time instead of a block of returns, we generate more type I errors

¹⁵The FWER dates back to [Bonferroni \(1936\)](#), and is defined as the probability of at least one false discovery. [Romano and Wolf \(2005a,b\)](#) introduce a stepwise multiple testing procedure that not only has higher statistical power than the tests of [Bonferroni \(1936\)](#) and [Holm \(1979\)](#) but also allows for cross-sectional dependence.

(false positive findings) than expected by the test. We also note that even if there is no serial dependence, our block-bootstrapped alpha test statistic is accurate.

4. Value-weighted Portfolios of Mutual Funds

As [Berk and Green \(2004\)](#) argue, funds managed by skilled managers attract greater portfolio flows than funds managed by unskilled managers. Hence, if we want to measure the skill of a fund manager, or if we want to test whether active investing is a zero-sum game, we must measure performance on a value-weighted basis and against an investable benchmark.

4.1 Alpha Relative to Investable Benchmarks

To compare the performance of active and index investing, we use investable benchmark models based on Morningstar’s investment categories, as described in Section 2.1. In Figure 1, we plot the robust p -values against the net and gross alphas for each of the available investment categories. As argued in Section 3, we adjust the p -values for multiple hypothesis testing using the method of [Romano and Wolf \(2016\)](#).

[Figure 1 about here.]

After fees, we find the “US Equity Large Cap Blend” category for institutional funds and the “Canada Fixed Income” category for retail funds to significantly underperform the alternative of the value-weighted index funds for both periods. For the negative alpha of the “Euro Fixed Income” retail category and the period from 1992 to 2016 we also find a significant p -value. Hence, only for three investment categories can we reject the zero-sum game hypothesis of [Berk and Green \(2004\)](#). Furthermore, our finding that “US Equity Large Cap Blend” institutional funds underperform after fees is perfectly in line with the argument of [Pastor, Stambaugh, and Taylor \(2015\)](#), in that higher competition in big active mutual fund industries leads to

diminishing returns to scale. Before fees, there are no investment categories with significantly negative alphas. However, we find “US Fixed Income” from 1992 to 2016, and “Global Equity Large Cap,” “Emerging Markets Equity,” and “Europe Equity Large Cap” from 2000 to 2016 for institutional, and also “Global Equity Large Cap” from 1992 to 2016 and “Global Equity” from 2000 to 2016 for retail funds to significantly outperform the value-weighted index funds.¹⁶

In addition to the hypothesis whether alpha is zero after fees, we can also ask how the value added by active management behaves over time and whether there are substantial differences across markets. In Figure 2, we aggregate the total value in US dollars that was extracted from the market by active management and compare it to the total fee value in US dollars that was paid by investors.¹⁷

[Figure 2 about here.]

For retail funds, Figure 2 tells us that active US funds fail to extract value from the market. For both equity and fixed-income funds, the value added is mostly close to zero. US equity funds experienced a severe drop during the second half of the 1990ies, recuperated in 2000 and the following years, but started to decline again after the recent financial crisis gradually. In contrast, US fixed income funds suffered a sudden drop during the financial crisis but profited considerably during the recovery from an increasing risk appetite and decreasing credit spreads.¹⁸

For active US equity funds, we find that the total extracted value before costs decreases since the financial crisis for both retail and institutional investors. While for institutional

¹⁶We remark that the choice of a block size of three is a conservative choice. As a robustness check, when we apply a block size of six or nine, the p -values increase slightly. The “Euro Fixed Income,” “Emerging Markets Equity,” and “Europe Equity Large Cap” categories, which all exhibit a p -value just below 10% for the block size of three, start to show insignificant p -values, further supporting the theory that alpha is zero after fees.

¹⁷In particular, we first compute the value-weighted gross alpha within an investment category against the value-weighted gross benchmark of the index funds and then aggregate the investment categories with the assets under management of all active funds within each category. We split into retail and institutional funds in the US and non-US investment categories.

¹⁸As we see in Table 2, active managers in the US show a significant exposure to the credit spread for both retail and institutional funds.

funds, the value extracted from the market becomes less than the value of the fees only recently in 2014, for retail funds the value added is only above the fees at the very beginning of our sample period. This observation, particularly for retail funds, resonates well with the negative value-weighted net alphas from our previous analysis and the poor performance of active management highlighted by [Fama and French \(2010\)](#) for the US equity market.

However, for non-US funds, [Figure 2](#) conveys a different message. Over the whole sample period, institutional non-US equity funds extract a value from the market that is substantially above their fees. Even for retail funds the value added is above their fees, at least during most of the sample period. In the fixed income space, while US institutional active managers extract more value compared to the total fees paid, non-US funds struggled recently. For both retail and institutional fixed income active managers, the total gross alpha decreased significantly for non-US investors around 2014. In a detailed analysis, we find that this decrease coincides with the beginning of quantitative easing in Europe and its impact on the European fixed income market. Active managers were underweighted in bonds of the periphery European countries that profited the most.¹⁹

We interpret our finding that non-US active equity funds add more value than US managers as supporting evidence to the theory of [Pastor, Stambaugh, and Taylor \(2015\)](#) in the sense that the higher competition in mature and large mutual fund industries decreases the fund’s ability to outperform passive benchmarks. Interestingly, the US fixed income market behaves differently, which may indicate that the competition in this segment remains low, regardless of fund size, potentially due to the higher complexity of the product.

Because the added values and fees in [Figure 2](#) depend on the total amount that was invested in each segment, we also analyze the evolution of the value-weighted alphas along the same dimensions as in [Figure 2](#). In [Figure 3](#), we show the cumulated aggregated alpha

¹⁹As we show later in [Table 2](#), active fixed income managers tend to have a negative duration exposure. Hence, the decreasing yields in European markets hurt their performance.

of active index funds over time.²⁰ Equal weighted, all investment categories for equity funds provide a positive alpha over time, even after fees. Value weighted, the aggregated alphas remain positive before fees, but they are zero for institutional funds and slightly negative for retail funds after fees. For the fixed income mutual funds, we find that the value-weighted alpha across investment regions is positive for the equal and value-weighted aggregation of the investment categories. For the institutional funds after fees, we also observe positive alphas over time. There are three major periods where active managers underperformed their index counterpart: equity funds before the burst of the dot-com bubble, both equity and fixed income funds in the financial crisis, and a slight underperformance in the recent past, especially after fees.

[Figure 3 about here.]

Garleanu and Pedersen (2018) argue that small investors tend to underperform because of their higher search costs and fees, while large investors are expected to outperform after a certain size because of their economies of scale and lower fees. For equity funds, the patterns observed in Figure 3 confirm their theory. We observe higher aggregated alphas for institutional funds after costs, but similar alphas for retail equity managers before costs. Since retail funds can pool the investments of small investors, and mutual funds often manage retail and institutional money in the same aggregated fund, we expect this pattern. For fixed income funds, we make a similar observation.

What is surprising in Figure 3, however, is the existence of three major periods where active managers underperformed their index counterparts: before the burst of the dot-com bubble, in the financial crisis, and in the recent past. Since we would have expected that

²⁰We first compute the value-weighted alpha within an investment category against the value-weighted benchmark of the index funds and then aggregate the investment categories with equal and value-weights. We also split into retail and institutional funds before and after fees.

For the equity mutual funds, we find that active mutual funds provide a superior alpha than index funds in every analysis before fees.

active management pays in turbulent times, we will further explore this observation in the next section.

4.2 Drivers for the Difference in Performance of Active and Index funds

To gain further intuition about what drives a wedge between the performance of the average active fund and that of the average index fund, we ask whether the multi-factor model provides some explanation for the difference in returns between the value-weighted portfolios of active and index mutual funds. Alerted by our observation from Figure 2, we enrich our regressions with the volatility index (VIX) of the Chicago Board Options Exchange (CBOE) as a fear gauge to proxy for market uncertainty.²¹

Table 2 shows the results for both active equity and fixed income funds when measured against the investable benchmark. Overall, the average values of R^2 absorb a significant fraction of the variance of the alpha before fees, in particular for fixed income funds and US institutional equity funds. For the equity funds in Panel A, we observe that the gross alpha loads profoundly and significantly on the SMB factor. Also, especially in the US, the performance difference loads negatively on the HML factor. The exposure is more pronounced for retail funds. Institutional funds, in contrast, have a much lower exposure to growth stocks. Overall, active funds seem to have a prominent small-cap bias and favor growth over value stocks. Furthermore, they tend to load positively on the momentum factor and negatively on the betting-against-beta factor.

Furthermore, we observe that the US equity market is the only market that shows a negative exposure to value (HML) for both retail and institutional investors. This finding implies that active managers favor glamour stocks as highlighted by Gennaioli, Shleifer, and Vishny

²¹We downloaded the time-series of the VIX index of the Chicago Board Options Exchange (CBOE) from Bloomberg. In unreported results, we find that a high level of proxied uncertainty, e.g., by earnings-per-share volatility or dispersion of returns within a fund category, is in general favorable for the performance of actively managed funds. However, these effects are not significant.

(2015) to attract investors to invest in their funds in this highly competitive mutual fund industry. The benefit from pandering might be more important compared to the higher return of contrarian strategies and might also be an explanation for the lower overall performance of active managers in the highly competitive US market.

[Table 2 about here.]

Concerning the VIX, we find that the difference in performance of active and index investing shows a negative sensitivity to changes in the VIX, which is often statistically significant. At first sight, this finding seems to run against our intuition, as we would expect active managers to use their skill to anticipate sudden uncertainty shocks. However, active managers that protect their portfolio against adverse shocks must pay an insurance premium in the long term. Such protection would generate relative losses to the market return in good times. Therefore, our result suggests that active managers prefer to run a short exposure to general market volatility, i.e., they tend to prefer small gains by selling insurance.

In Table 2, Panel B, we see that fixed income managers have a negative exposure to the shift factor. Consequently, they are less affected by rising interest rates. In exchange, they load on other risk factors to compensate for the lower expected returns. In particular, they load significantly on the credit risk factor. As in Panel A for equity funds, fixed income managers also have a negative exposure to changes in the VIX. Therefore, they lose money if the VIX increases sharply, as it did during the latest financial crises, which also explains the large drop in the cumulative alpha in Figure 2 towards the end of 2008.

4.3 Persistence Analysis

From an investor’s perspective, it may be disappointing that active fund investing provides, by and large, zero alpha after fees. How then can a fund investor do better and profit from actively managed funds? An initial idea is provided by Carhart (1997). He finds that US equity mutual

funds with a substantial underperformance over the past year persist to underperform over the next year relative to a multi-factor benchmark. In contrast to the outperformance of the best mutual funds, he cannot explain the persistence in the worst mutual funds. Thus, it would be of interest to know whether this observation carries over to our setting with investable benchmarks.

To simulate the returns to an average active investor who trades according to this simple rule, we build momentum portfolios of active funds as follows. Every year in December, we first sort the active funds within each investment category based on their t -value for the value-weighted alpha measured against the investable benchmark.²² Then, we invest in the value-weighted portfolio of the $x\%$ best performing active funds and normalize the weights each month. We repeat the same exercise for the $x\%$ worst performing funds. If one month there is no data for a particular fund, it disappears from the portfolio. To aggregate the performance numbers of the different investment categories, we value-weight the net returns by the total active assets.²³ We assume that funds do not charge transaction costs for incoming and outgoing investors.

[Table 3 about here.]

Table 3 presents the momentum portfolio returns when selecting the best performing (Panel A) and the worst performing funds (Panel B). In Panel A, we find that the performance increases the more we focus on the best performing funds. For equity funds, the alpha after fees climbs from -0.23% (-0.60%) to a remarkable 0.62% (-0.02%) for institutional (retail) funds. For retail fixed income funds, alpha increases from -0.75% to -0.40% , but decreases for

²²To compute the alpha, we require at least ten of the twelve most recent monthly returns. When we sort by the t -value for the alpha, we consider the market risk of the fund and look at both the relative performance and the consistency of the relative performance against the benchmark.

²³For $x\%$ we chose steps of 10% starting with all mutual funds to the best 10% mutual funds. We disregard data points where we have less than ten active mutual funds, and to calculate the benchmark return for the alpha we must have at least one index fund within the category for the past twelve months to start the out-of-sample backtest. For some small investment categories, over the year the number of funds drops below ten. In this case, we apply the next less restrictive filter.

institutional fixed income funds from 0.26 to 0.16.²⁴ However, all these results are statistically insignificant, as the robust p -values remain high or even increase. Hence, after fees, even the best performing funds provide an alpha for the fund investor, which is statistically not different from zero. This absence of persistence supports the theoretical argument of [Berk and Green \(2004\)](#) that persistence should not exist since new money flows into well-performing funds and there are diseconomies of scale, or because successful funds capture excess returns by raising fees.²⁵

In Panel B of Table 3, we form portfolios by selecting the $x\%$ worst performing funds. We find that the value-weighted performance decreases drastically. For instance, for institutional equity funds it drops from -0.23% to -0.94%. However, only for retail funds does the negative performance of the 10% worst performing funds survive our robust alpha test adjusted for multiple hypothesis testing. For equity retail funds, the performance drops to -1.39% at the 5% significance level and, for fixed income retail funds, it drops to -0.84% at the 10% significance level. Hence, while [Carhart \(1997\)](#) shows for US equity funds and under a multi-factor benchmark that the persistence is significant for the worst performing funds, we can confirm this result only for retail equity and fixed income funds. Obviously, from these funds, the investor receives a negative alpha after costs. For institutional funds, we do not find such evidence. At the same time, Table 3 confirms [Carhart \(1997\)](#) in that we do not find any unexplainable persistence in overperforming funds.

[Chen et al. \(2004\)](#) and [Yan \(2008\)](#) find a negative relation between alpha and size and a positive relation with past return. Thus, on average, they find that future alpha is smaller for large funds but past returns are associated with higher future alpha, and predictability exists. Recently, [Elton, Gruber, and Blake \(2012\)](#) show that alpha persistence does not disappear for

²⁴Interestingly, for fixed income institutional funds, we observe substantially lower betas the more we exclude badly performing funds from the portfolio, indicating that the best performing fixed income funds run a slightly different exposure than their investable benchmark suggests.

²⁵As an additional exercise, we also explored the persistence of gross alphas. In unreported results, we find that the alphas increase substantially, but still they do not survive our statistical test, not even for the 10% best performing funds.

larger funds. To explore the interplay between size and predictability, we perform a bivariate sort on size and persistence.

[Table 4 about here.]

Table 8 presents the alphas after fees for the different portfolios sorted according to fund size and the previous year’s performance. We find that, except for fixed income retail funds, small winner funds perform better than their larger counterparts. They produce the highest alphas relative to their investable benchmarks. However, this outperformance is statistically insignificant. At the same time, we identify the small loser funds as the funds with the worst performance. For retail equity and fixed income funds, the negative alphas of small (and medium) loser funds become even statistically significant. Unsuccessful and small funds will continue to be unsuccessful, and they do so in a statistically significant way, while large funds tend to underperform but not significantly so. These findings resonate well with [Berk and Green \(2004\)](#)’s hypothesis that fund performance is inversely related to size due to diseconomies of scale.

4.4 Is the US Active Fund Industry Oversized?

As we observe in [Figure 1](#), “Europe Equity Large Cap Inst.” provides a significantly positive value-weighted gross alpha, while “US Equity Large Cap Blend Inst.” generates a significantly negative net alpha. This observation, together with the existence of diseconomies of scales, leads us to suspect that the disappointing performance of US equity funds relative to other equity funds is due to a potentially overfunded US industry. To shed light on this issue, we apply the methodology of [Zhu \(2018\)](#) to US and European equity funds. In particular, we assume that in a simplified setting the gross alpha α_g depends on fund size q as

$$\alpha_g = a - b \log(q). \tag{1}$$

If the manager optimizes the value added, defined as $V = \alpha_g q$, the optimal actively managed amount is just $q^* = \exp(a/b - 1)$. Moreover, the critical fund size at which the value added becomes zero is $q^c = \exp(a/b)$.

We can classify a fund i if $q_i < q_i^*$ as underfunded, as moderately overfunded if $q_i \in (q_i^*, q_i^c]$, and as excessively overfunded if $q_i > q_i^c$. To calculate the optimal and critical funds sizes, we estimate the b_i 's and a_i 's as in [Zhu \(2018\)](#), i.e., to minimize the effect of estimation errors, we first estimate b by forming decile portfolios, and then we estimate the individual a_i 's. Also, we use the mean size over a fund's lifetime as a proxy for q_i . Equipped with these estimates, we can then calculate the fractions of over- and underfunded funds.

[Figure 4 about here.]

Figure 4 confirms our suspicion for the US market. Over the whole sample period, we find a fraction of 52% of excessively overfunded US equity funds, which aligns well with the 54% found in [Zhu \(2018\)](#).²⁶ Furthermore, 20% are moderately overfunded, and only 28% of the funds are underfunded. We further observe a trend in these fractions. The number of excessively overfunded funds increases from 33% in 1992 to 59% by the end of 2016, while the underfunded fraction decreases from 54% to 22%. When analyzing the fraction of AuM managed by excessively overfunded managers, we find that this number increased from 29% to 58%; hence, this share has doubled over the sample period.

In Figure 4, we also present the corresponding plots for the European funds. In contrast to the US, we identify over the whole sample only 32% as excessively overfunded and 53% as underfunded. As for the US, we find an increase in the number of overfunded funds over time, but this increase is less pronounced than in the US and at a lower level. The fraction of underfunded funds decreases from 66% in 1992 to 56% by the end of 2016. The excessively overfunded mutual funds gain in importance with a share of 19% at the beginning and 29% by

²⁶Our US fund data consist of 6,601 funds with a minimum size of 5 million December 2016 US dollars in assets under management (AuM), which have at least a history of three years of monthly data. [Zhu \(2018\)](#) analysis includes only 3,077 funds.

the end of our sample period. While the fraction of AuM managed by excessively overfunded funds has doubled in the US, for European funds this number has plunged from 66% to 21% over the same period.

Figure 4 also shows the immense growth of the mutual fund industry, both in the US as well as in Europe. However, the growth in the US has been much stronger. The average monthly AuM increased from 1992 to 2016 by a factor of 90, whereas the AuM in Europe increased by a factor of 31. At the same time, the AuM per fund increased by a factor of 7.2 and only by a factor of 1.12 in Europe. At the backdrop of these numbers, we provide convincing evidence that the disappointing performance of the US mutual fund industry over the recent years is, in no small part, a manifestation of diseconomies of scale.²⁷ Moreover, given the trends in the fraction of AuM managed by excessively overfunded funds in the US and Europe, as observed in Figure 4, the US fund industry may come under increasing pressure.

4.5 Impact of Fees

For US equity funds, [Gil-Bazo and Ruiz-Verdú \(2009\)](#) find a puzzling underperformance of mutual funds that charge higher fees. Their finding contradicts the argument of [Habib and Johnsen \(2016\)](#) that higher fees act as a signal for the unobservable quality of the costly research by active managers. However, higher fees can also be seen as a sure loss for investors, since they directly reduce the portfolio return when the quality of the manager is unobservable. Thus, higher fees imply lower net returns if the costly research of the active manager does not improve performance. To shed more light on this debate, we explore the impact of fees by proceeding analogously to the persistence analysis of the previous section.

[Table 5 about here.]

²⁷We also did the same analysis just for “Europe Equity Large Cap Inst.” and “US Equity Large Cap Blend Inst.” For these two categories, the differences are even more striking. For this European investment category, we find that over the whole sample period 72% of AuM is managed by underfunded funds, and in December 2016, this number was as high as 82%. The corresponding numbers for the US investment category are 24% and 23%, respectively, and excessively overfunded funds managed 60% of AuM by the end of 2016.

Using an investable benchmark, Table 5 presents the performance of active fund portfolios that include the $x\%$ least expensive (Panel A) and the $x\%$ most expensive (Panel B) funds of the preceding year. In Panel A, the alpha against the investable benchmark increases for institutional equity funds from -0.23% to 0.83% and from -0.60% (-0.75%) to 0.30% (-0.40%) for retail equity (fixed income) funds. Hence, even the alpha of retail equity funds gets into positive territory if we exclude those funds that charge the highest fees. Again, although a fund investor can improve the performance in their fund portfolio by including only the least expensive funds, these improvements are statistically insignificant. For institutional fixed income funds, the alpha first increases but then decreases for the fund portfolio with the 20% and 10% lowest fees. Interestingly, the betas of these portfolios are substantially below one.

Panel B of Table 5 shows the performance of the portfolios with the $x\%$ most expensive funds. If the argument of [Habib and Johnsen \(2016\)](#) were valid and high fees were a signal of quality, we would expect increasing alphas the more we filter out the cheaper funds. However, we observe the opposite. The portfolios with the 10% most expensive funds perform poorly. For instance, the performance of the portfolio of the 10% institutional equity funds drops to -1.14% , compared to the 10% cheapest fund with an alpha of 0.83% . For equity retail funds, the underperformance of high-fee funds becomes significant already when we look at the 90% most expensive funds. For the 10% most expensive retail funds, we get a highly significant alpha of -2.83% , compared to the 10% least expensive retail funds with an alpha of 0.30% . Hence, if we only consider the universe of the most expensive equity retail funds, active funds are detrimental to retail investors.

[Table 6 about here.]

Given the evidence that past winners and low fee funds generate a higher average alpha over time, we next ask whether the same pattern emerges when we control for performance persistence and fees simultaneously. Hence, we build nine portfolios that arise from the bi-

variate sort and the 30th and 70th percentiles for each criterion.²⁸ Table 6 shows the results. All active alphas are positive for the low-fee and winner portfolios except for the fixed income retail investor. Furthermore, for all sorted portfolios, the alpha in the top right (low fee and winner) corner is always larger than the alpha in the lower left (high fee and loser) corner. These alphas are insignificant, except for the equity retail funds. Here, the high-fee and loser portfolio has a highly significant negative alpha. Overall, we find that both high performance and low fees over the past year have a positive impact on the alpha in the next year. The result is robust in the bivariate sort of the two criteria. Especially for retail funds, an investor is well advised to avoid high-fee loser funds as the negative alphas are highly significant under our testing framework.

[Figure 5 about here.]

To shed further light on which funds charge higher fees, Figure 5 shows the average active fees of the highly competitive US equity market over time together with the relative share of index funds in terms of assets under management. As expected, we find a substantial difference between the fees charged by retail and institutional funds, depending on the fund's age. Young retail funds charge the highest fees. However, the size of their fees has drastically decreased since the recent financial crisis, converging to the level of the fees charged by older retail funds.

Interestingly, over the whole period, young institutional funds have charged fees similar to their older competitors. The gap between old retail and old institutional funds has been somewhat steady over the years, slightly narrowing recently. We also find that expense ratios are lower for large funds. By the end of 2016, we find that the average expense ratio of the equal-weighted portfolio compared to the value-weighted portfolio is 21% higher for institutional equity funds, 37% higher for retail equity funds, 17% higher for institutional

²⁸For some small investment categories, there are time periods where none of the mutual funds belong to a particular group. In such a case, we invest in the value-weighted portfolio of all active funds within this category.

fixed-income funds, and 34% higher for retail fixed-income funds. Since institutional funds are usually larger than retail funds, this gap is consistent with the finding of [Elton, Gruber, and Blake \(2012\)](#) that expense ratios are lower for larger funds. As [Figure 5](#) suggests, over the years, the level of fees for active funds has tended to decrease further. Thus, we find further evidence for the zero-sum game hypothesis of [Berk and Green \(2004\)](#) in the sense that active managers start to adjust their fees due to the unabated growth of index funds.

5. Conclusion

Analyzing a rich dataset from Morningstar covering 61,269 mutual funds from different regions and asset classes from 1992 to 2016 and comparing their returns to investable benchmarks, which we construct as portfolios of index funds. We find significant negative alphas after fees only for the “US Equity Large Cap Blend” for institutional funds and “Canada Fixed Income” for retail funds. For the vast majority of categories, we cannot reject the null hypothesis that net alphas to investors are zero. Indeed, we even find categories such as “US Fixed Income” and “Global Equity Large Cap” for institutional investors with significant p -values before fees.

At first glance, one would expect active managers to invest more carefully and take fewer risks. We have confirmed this hypothesis by the fact that active management takes a more conservative position to the traditional risk factors, such as market and duration risk. However, we find that active equity and fixed income mutual funds are affected by adverse volatility shocks, suggesting that active managers sell protection to collect the insurance premium. Also, averaging over the different regions, we find that the active investor has a higher sensitivity than index funds to alternative risk premia such as small cap and credit risk.

Sorting active fund portfolios according to their performance persistence, fees, and size, we find that low-fee winner portfolios and small winner portfolios tend to outperform, but their alpha does not survive our test statistics. These results give further support to the

competitive equilibrium proposed by [Berk and Green \(2004\)](#). Our analysis also highlights some substantial differences between institutional and retail funds. In particular, our empirical results suggest to active retail investors that they should avoid high-fee losers and small losers. Their alphas are negative and statistically significant, surviving our robust test statistics adjusted for multiple hypotheses.

References

- Agarwal, V., and Naik, N. Y. (2004) Risks and portfolio decisions involving hedge funds, *The Review of Financial Studies* 17, 63–98.
- Andrikogiannopoulou, A., and Papakonstantinou, F. (2016) Estimating mutual fund skill: A new approach, *Swiss Finance Institute Research Paper* .
- Andrikogiannopoulou, A., and Papakonstantinou, F. (2018) Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power?, *Journal of Finance* forthcoming.
- Banegas, A., Gillen, B., Timmermann, A., and Wermers, R. (2013) The cross section of conditional mutual fund performance in European stock markets, *Journal of Financial Economics* 108, 699–726.
- Barras, L., Scalliet, O., and Wermers, R. (2010) False discoveries in mutual fund performance: Measuring luck in estimated alphas, *Journal of Finance* 65, 179–216.
- Berk, J. B., and Green, R. C. (2004) Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.
- Berk, J. B., and van Binsbergen, J. H. (2015) Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.
- Berk, J. B., and van Binsbergen, J. H. (2017) Mutual funds in equilibrium, *Annual Review of Financial Economics* 9, 147–167.
- Bonferroni, C. E. (1936) *Teoria statistica delle classi e calcolo delle probabilita* (Libreria Internazionale Seeber).
- Carhart, M. M. (1997) On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

- Chen, J., Hong, H., Huang, M., and Kubik, J. D. (2004) Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* 94, 1276–1302.
- Crane, A. D., and Crotty, K. (2018) Passive versus active fund performance: Do index funds have skill?, *Journal of Financial and Quantitative Analysis* 53, 33–64.
- Elton, E. J. (2001) A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases, *Journal of Finance* 56, 2415–2430.
- Elton, E. J., Gruber, M. J., and Blake, C. R. (2012) Does mutual fund size matter? The relationship between size and performance, *The Review of Asset Pricing Studies* 2, 31–55.
- Fama, E. F., and French, K. R. (1992) The cross-section of expected stock returns, *Journal of Finance* 47, 427.
- Fama, E. F., and French, K. R. (2010) Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915–1947.
- Ferreira, M. A., Keswani, A., Miguel, A. F., and Ramos, S. B. (2013) The determinants of mutual fund performance: A cross-country study, *Review of Finance* 17, 483–525.
- Frazzini, A., and Pedersen, L. H. (2014) Betting against beta, *Journal of Financial Economics* 111, 1–25.
- French, K. R. (2008) Presidential address: The cost of active investing, *Journal of Finance* 63, 1537–1573.
- Garleanu, N. B., and Pedersen, L. H. (2018) Efficiently inefficient markets for assets and asset management, *Journal of Finance* forthcoming.

- Genest, C., and Rémillard, B. (2004) Test of independence and randomness based on the empirical copula process, *Test* 13, 335–369.
- Gennaioli, N., Shleifer, A., and Vishny, R. (2015) Money doctors, *Journal of Finance* 70, 91–114.
- Gerakos, J., Linnainmaa, J. T., and Morse, A. (2016) Asset managers: Institutional performance and smart betas, NBER Working Papers 22982, National Bureau of Economic Research.
- Gil-Bazo, J., and Ruiz-Verdú, P. (2009) The relation between price and performance in the mutual fund industry, *The Journal of Finance* 64, 2153–2183.
- Habib, M. A., and Johnsen, D. B. (2016) The quality-assuring role of mutual fund advisory fees, *International Review of Law and Economics* 46, 1–19.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics* 65–70.
- Jegadeesh, N., and Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65.
- Kosowski, R. (2011) Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions, *The Quarterly Journal of Finance* 1, 607–664.
- Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006) Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* 61, 2551–2595.
- Kuensch, H. R., and Goetze, F. (1996) Second-order correctness of the blockwise bootstrap for stationary observations, *The Annals of Statistics* 24, 1914–1933.

- Ledoit, O., and Wolf, M. (2008) Robust performance hypothesis testing with the Sharpe ratio, *Journal of Empirical Finance* 15, 850–859.
- Ledoit, O., and Wolf, M. (2011) Robust performances hypothesis testing with the variance, *Wilmott* 2011, 86–89.
- Moskowitz, T. J. (2000) Discussion, *The Journal of Finance* 55, 1695–1703.
- Pastor, L., Stambaugh, R. F., and Taylor, L. A. (2015) Scale and skill in active management, *Journal of Financial Economics* 116, 23 – 45.
- Patton, A., Politis, D. N., and White, H. (2009) Correction to “Automatic block-length selection for the dependent bootstrap” by D. Politis and H. White, *Econometric Reviews* 28, 372–375.
- Pedersen, L. H. (2018) Sharpening the arithmetic of active management, *Financial Analysts Journal* 74, 21–36.
- Pesaran, H. M. (2004) General diagnostic tests for cross section dependence in panels, CESifo Working Paper Series 1229, CESifo Group, Munich, Germany.
- Politis, D. N., and White, H. (2004) Automatic block-length selection for the dependent bootstrap, *Econometric Reviews* 23, 53–70.
- Ratcliff, J. W., and Metzener, D. E. (1988) Pattern-matching: The gestalt approach, *Dr Dobbs Journal* 13, 46.
- Romano, J. P., and Wolf, M. (2005a) Exact and approximate stepdown methods for multiple hypothesis testing, *Journal of the American Statistical Association* 100, 94–108.
- Romano, J. P., and Wolf, M. (2005b) Stepwise multiple testing as formalized data snooping, *Econometrica* 73, 1237–1282.

- Romano, J. P., and Wolf, M. (2016) Efficient computation of adjusted p -values for resampling-based stepdown multiple testing, *Statistics & Probability Letters* 113, 38–40.
- Sharpe, W. F. (1991) The arithmetic of active management, *Financial Analysts Journal* 47, 7–9.
- van Binsbergen, J. H., and Opp, C. C. (2019) Real anomalies, *The Journal of Finance* forthcoming.
- Wermers, R. (1999) Mutual fund herding and the impact on stock prices, *The Journal of Finance* 54, 581–622.
- Yan, X. S. (2008) Liquidity, investment style, and the relation between fund size and fund performance, *Journal of Financial and Quantitative Analysis* 43, 741–767.
- Zhu, M. (2018) Informative fund size, managerial skill, and investor rationality, *Journal of Financial Economics* 130, 114–134.

A. Description of the Data

We summarize the steps for the data cleaning of the Morningstar database and provide summary statistics for the different asset classes and investment categories.

A.1 Raw Morningstar Data

Our mutual fund sample is from the Morningstar database. Recent work in [Kosowski et al. \(2006\)](#); [Fama and French \(2010\)](#); [Barras, Scalliet, and Wermers \(2010\)](#) concentrates mostly on the survivor-bias-free CRSP US Mutual Fund Database. As shown by [Elton \(2001\)](#), the CRSP database also suffers from a survivorship bias: the so-called omission bias. [Berk and van Binsbergen \(2015\)](#) find that neither the CRSP nor the Morningstar database are free from errors. Thus, we must be careful, and we find the same errors as reported in this previous paper. We focus on all funds with an *Investment Type* flagged by “Open-End Fund” or “Exchange-Traded Fund” including non-survivors from December 1991 to December 2016. We downloaded the following fields for each share class.

For the description of a share class, we retrieved the *Name*, *ISIN*, and *Base Currency*. It is common to name a share class starting with the name of the asset manager, followed by a description of the strategy, and an ending for the share class. For example, for the equity fund “Blackrock S&P 500 Index,” there is a share class “Blackrock S&P 500 Index Institutional” for institutional and the “Blackrock S&P 500 Index Investor A” for retail clients.

The most specific categorization in Morningstar is the *Morningstar Category*, which is derived by analyzing the underlying portfolio holdings. In all, we find 504 different groups for the retail equity and fixed income funds. The *Global Category* combines several Morningstar categories, and we see a total of 68 groups for retail equity and fixed income funds. For example, the *Global Category* category “Europe Equity Large Cap” includes Morningstar categories, such as “EAA Fund Europe Large-Cap Blend Equity,” “EAA Fund Europe Large-Cap Value

Equity,” “EAA Fund Europe Large-Cap Growth Equity,” but also “US Fund Europe Stock” or “Canada Fund European Equity.” Since we have within this broader categorization a higher chance of finding both index and equity funds, we concentrate on the *Global Category*. The *Global Broad Category Group* further aggregates the *Global Category* into the major asset classes. Since we focus on the comparison of active and index funds, we concentrate on the *Global Broad Category Group* “Equity” and “Fixed Income” funds. We thereby disregard categories such as “Allocation,” “Money Market,” or “Commodities” because for them there are insufficiently many index funds to make a fair comparison.

For the computation of the returns, we downloaded the following fields for each fund: *Monthly Return USD*, *Monthly Gross Return USD*, and *Net Assets - share class (Monthly) USD*. The *Monthly Return USD* includes management, administrative, and other costs that are deducted from the NAV, such as the 12b-1 fee. All income and capital gains are reinvested monthly. The *Monthly Gross Return USD* is based on the *Monthly Return USD* and adds the most recent net expense ratio. The *Net Assets - share class (Monthly) USD* is the monthly total net assets of a share class.

To distinguish between active and index funds, we make use of the *Index Fund* field. Those funds that track a particular index based on full replication or based on a representative sampling are flagged by Morningstar as index funds. Next, to filter the institutional and retail funds, we downloaded the field *Institutional*, which defines any fund as institutional if it either says “institutional” in the name of its share class, has a minimum investment above USD 100,000, or the prospectus says that it is for institutional investors only.

A.2 Data Cleaning

For each fund, we retrieved its monthly net return, gross return, and total net assets, all in US dollars. We only included an observation if all three items were available. Often, and as reported in [Berk and van Binsbergen \(2015\)](#), we observe that net assets are reported quarterly

or are missing for a specific month. In this case, we roll the assets under the assumptions of zero net flows, so as to increase the available data points and avoid disconnected time-series. Besides, for some institutional mutual funds, we observe zero fees because they are paid in separate contracts with the asset manager. Thus, we only include funds where the sum of the gross returns is larger than the sum of the net returns to exclude zero-fees funds. To avoid the incubation bias, we include funds only if they reach 5 million December 2016 US dollars in AUM.

We also see conversion errors, where funds assets suddenly increase by a high factor and then decrease again by a similar factor. First, we observe this behavior in emerging market currencies before 1999. Thus, we concentrate in the period before 1999 only on the developed currencies, Pound Sterling, US Dollar, Euro, Singapore Dollar, Australian Dollar, Swedish Krona, South African Rand, Swiss Franc, Japanese Yen, New Zealand Dollar, Canadian Dollar, Norwegian Krone, Danish Krone. Also, we see that for some funds, the assets change by a factor higher than 100 and decrease in the next period to the same level as before the outlier. For these cases, we smooth the net assets over time if we see that the assets change by a factor higher than 10 and we decrease them in the next two periods by a factor of more than 0.5. But there are funds where this increase is verified by attaining the same fund levels in the future. Therefore, we only correct the assets if the same level is not exceeded in its future assets.

We also delete obvious mistakes, such as when an index fund shows high fees in the past and suddenly changes to a low fee. In this case, we keep only the low fee period, since we interpret this as being that either the fees were not correct or the fund changed from active to index. When we build the value-weighted portfolio for the investment categories, we also remove funds that show a beta below 0.05 relative to the average return of all the funds within the same investment category. Because of the low sensitivity to the average fund, these funds are not following a strategy similar to that of the rest of the group.

A.3 Aggregation of the Share Classes

Each line in the Morningstar dataset corresponds to a share class. In all, we obtain 435,453 lines of different share classes. Thus, we must aggregate the same share classes to avoid multiple tries of investment strategies by the same provider. First, we tried to use the fields *Administrator* and *Ticker of Fund's Oldest Share Class*; however, they are often missing. For this reason, we aggregated alphabetically subsequent mutual funds that are in the same Morningstar Category with the corresponding *Index Fund* flag and have a similar name. While [Berk and van Binsbergen \(2015\)](#) use the last word of the fund's name for the share class, we use the ratio provided by the SequenceMatcher of the difflib library in Python, which is based on the algorithm developed by [Ratcliff and Metzener \(1988\)](#) and, additionally, cleans the “junk” elements. We define two names to be similar if this ratio is above 0.8.

A.4 Summary Statistics of Investment Categories

Table 1 shows the summary statistics of the cross-sectional monthly attributes across asset classes. Table A.1 provides a more detailed view of all the investment categories, where we find both index and active mutual funds. For the active funds, we analyze a total of 14,969 institutional and 46,300 retail funds, of which 56,136 are active funds and 5,133 are index funds. In general, there are fewer index funds, but they have higher average total net assets (TNA) and net returns, and also lower fees and about the same average number of years in the database. As expected, the institutional funds charge lower fees than their retail counterparts.

[Table A.1 about here.]

B. The Robust Alpha Test

Consider a fund with time- t return y_t and a set of K benchmark factor returns x_{tk} , $k = 1, \dots, K$. A total of T returns are observed. We assume that these observations are generated by a stationary multivariate return distribution with mean vector μ and covariance matrix Σ :

$$\mu = \begin{pmatrix} \mu_y \\ \mu_{x_1} \\ \vdots \\ \mu_{x_K} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{yx_1} & \cdots & \sigma_{yx_K} \\ \sigma_{x_1y} & \sigma_{x_1}^2 & \cdots & \sigma_{x_1x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_Ky} & \sigma_{x_Kx_1} & \cdots & \sigma_{x_K}^2 \end{pmatrix}, \quad (\text{B.1})$$

with the observed means $\hat{\mu}$ and sample covariance matrix $\hat{\Sigma}$. By defining a vector $\mu_X = (0, E[x_1], \dots, E[x_K])'$, we can express the fund's alpha as

$$\alpha = E[y] - \mu_X' \Sigma_{XX}^{-1} y_X, \quad (\text{B.2})$$

with

$$\Sigma_{XX} = \begin{pmatrix} 1 & E[x_1] & E[x_2] & \cdots & E[x_K] \\ E[x_1] & E[x_1^2] & E[x_1x_2] & \cdots & E[x_1x_K] \\ \vdots & \vdots & & \ddots & \vdots \\ E[x_K] & E[x_Kx_1] & E[x_Kx_2] & \cdots & E[x_K^2] \end{pmatrix} \quad \text{and} \quad y_X = \begin{pmatrix} E[y] \\ E[x_1y] \\ \vdots \\ E[x_Ky] \end{pmatrix}. \quad (\text{B.3})$$

Then, we test for the hypothesis

$$H_0 : \quad \alpha = 0 \quad \text{against} \quad H_1 : \quad \alpha \neq 0. \quad (\text{B.4})$$

Furthermore, we define $\zeta_k = E[yx_k]$, $\gamma_k = E[x_k^2]$, $\xi_{kj} = E[x_k x_j]$, $j > k$, and the combined vector $\nu = (\mu_y, \dots, \mu_{x_k}, \dots, \zeta_k, \dots, \gamma_k, \dots, \xi_{kj}, \dots)' \in \mathbb{R}^{1+3k+k(k-1)/2}$ with sample counterpart $\hat{\nu}$. Now, we can express the true alpha as a function f of ν :

$$\alpha = E[y] - \mu'_X \Sigma_{XX}^{-1} y_X = f(\nu); \quad (\text{B.5})$$

and the estimated alpha as function of $\hat{\nu}$: $\hat{\alpha} = f(\hat{\nu})$. As mentioned in [Ledoit and Wolf \(2008\)](#), under mild regularity conditions,

$$\sqrt{T}(\hat{\nu} - \nu) \xrightarrow{d} N(0, \Psi), \quad (\text{B.6})$$

where Ψ is an unknown symmetric positive semi-definite matrix. By the delta method, we obtain

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \nabla' f(\nu) \Psi \nabla f(\nu)) \quad (\text{B.7})$$

with

$$\nabla' f(\nu) = \left(\frac{\partial f(\nu)}{\partial \mu_y}, \dots, \frac{\partial f(\nu)}{\partial \mu_{x_k}}, \dots, \frac{\partial f(\nu)}{\partial \zeta_k}, \dots, \frac{\partial f(\nu)}{\partial \gamma_k}, \dots, \frac{\partial f(\nu)}{\partial \xi_{kj}}, \dots \right)' \quad (\text{B.8})$$

Given a consistent estimator $\hat{\Psi}$ of Ψ , we can compute a standard error for $\hat{\alpha}$ by

$$s(\hat{\alpha}) = \sqrt{\frac{\nabla' f(\nu) \hat{\Psi} \nabla f(\nu)}{T}}. \quad (\text{B.9})$$

To test the null hypothesis in Equation (B.4), we focus on the bootstrap inference for time-series data outlined in [Ledoit and Wolf \(2008\)](#). In particular, we denote the optimal block length by b and define $l = \text{floor}(T/b)$. As shown in [Kuensch and Goetze \(1996\)](#), the bootstrapped estimator of $\hat{\Psi}^*$ is

$$\hat{\Psi}^* = \frac{1}{l} \sum_{j=1}^l \eta_j \eta_j', \quad (\text{B.10})$$

where

$$z_t^* = \left(y_t^* - \hat{\mu}_y^*, \dots, x_{tk}^* - \hat{\mu}_x^*, \dots, y_t x_{tk} - \hat{\zeta}_k^*, \dots, x_{tk}^2 - \hat{\gamma}_k^*, \dots, x_{tk}^* x_{tj}^* - \hat{\xi}_{kj}, \dots \right),$$

$$\eta_j = \frac{1}{\sqrt{b}} \sum_{t=1}^b z_{(j-1)b+t}^*.$$
(B.11)

Next, the studentized statistics are

$$\tilde{d}_m^* = \frac{|\hat{\alpha}_m^* - \hat{\alpha}|}{s(\hat{\alpha}_m^*)},$$
(B.12)

and the p -value is

$$PV = \frac{\{\tilde{d}_m^* \geq \hat{d}\} + 1}{M + 1},$$
(B.13)

where \hat{d} is the original studentized test statistic that was computed from the observed returns. We use Newey–West standard errors to calculate the original standard errors. Regarding the optimal block, we suggest using either the method of [Politis and White \(2004\)](#) and [Patton, Politis, and White \(2009\)](#) for the univariate case, or the method of [Leford and Wolf \(2008\)](#) for the bivariate case. For our empirical analysis, we would like to compare up to 30 different investment categories and, so far, there is no available method to make this comparison. Consequently, we will further discuss the optimal block size to use in our simulations in [Appendix C](#).

C. Accuracy of the Robust Alpha Test

We now present the results of a simulation study to show the difference between our robust alpha test and the standard hypothesis tests. For this purpose, we first simulate a single hypothesis setting. For realistic time-series, we select the first ten US mutual funds of the Morningstar database within the category “US Equity Large Cap Blend” that offer the entire return history from 1992 to 2016 ($T=300$). As benchmark models, we focus on the one-factor

“CAPM,” i.e., the market excess return, the three-factor “FF3,” and the five-factor “FF5” model. For the data generating process (DGP), we sample from the realized returns with a circular block bootstrap and block sizes of 1, 3, and 6. We selected this grid of block sizes based on our analysis in Section 2, where we observe that most of the optimal circular block sizes range from one to six. This grid corresponds to time periods of one, three, and six months. The block sizes of three and six are the ones that take the evidence of serial dependence from Section 2 into account. A block size of one generates independent data, and we employ this block size only for reasons of comparison. For each fund, we simulate 1,000 paths and set the alpha under the null hypothesis to the true observed alpha of the data. The bootstrapped p -values (Boot) are then calculated as illustrated in Appendix B by employing $M = 1,000$ and the optimal block size by the method of Politis and White (2004) and the correction of Patton, Politis, and White (2009). We compare the robust p -values with those from the standard inference methods; that is, based on the normal distribution (Standard), Newey–West (NW), and HC3 standard errors.

[Table C.1 about here.]

Table C.1 shows the empirical rejection probabilities of the falsely rejected null hypothesis compared to the nominal levels $\alpha = 10\%$, $\alpha = 5\%$, and $\alpha = 1\%$. Because the null hypothesis is true for all the simulations, the true rejection probabilities should be equal to the nominal levels of the test. If a test shows a higher percentage of rejections, then we regard this test as too liberal. While we observe that the standard inference tests based on the normal distribution, the Newey–West, and HC3 standard errors are too liberal in rejecting the null hypothesis, the bootstrapped solution (Boot) presented in the previous section is close to the nominal levels. We highlight in bold the empirical rejection probabilities that are closest to the desired level. We observe the HC3 standard errors to be in some cases closer to the desired level than are those of the block bootstrapped method, but only in the case where we apply the standard but less realistic bootstrap with a block size of one where we lose any

dependence over time. However, as we demonstrate in Section 2, the optimal block size, and thus a realistic assumption for the DGP is, in general, around three or six, for which our bootstrapped test is tailored to be more accurate. A similar observation was also made in [Ledoit and Wolf \(2008\)](#) for testing the Sharpe Ratio and in [Ledoit and Wolf \(2011\)](#) for the variance.

Since there is still the open question of the optimal block size in the multiple hypothesis setting when controlling the FWER, as illustrated in [Romano and Wolf \(2005a,b, 2016\)](#), we conduct a second simulations study. To this end, we must consider the cross-dependence structure, and jointly sample the funds and benchmark returns. For this purpose, we focus on the 17 portfolios within the “Inv. Categories” setting from Section 2 with the investable one-factor benchmark model that is based on the value-weighted return of index funds. Also, instead of calculating the Type I Errors as in the single hypothesis setting, we compute the empirical rejection probabilities based on the FWER, as illustrated in [Romano and Wolf \(2005a,b, 2016\)](#). To find the optimal block size that is closest to the nominal levels of the test, we focus on the following grid of block sizes: 1, 3, 6, 9, and 12. Regarding the DGP, we keep the grid from our first simulation study.

[Table C.2 about here.]

Table C.2 shows the empirical rejection probabilities based on the FWER. Likewise, for the FWER, we find the bootstrapped robust alpha test to achieve the desired levels at optimal block sizes three or six. Given that for a block size of three we observe accurate rejection probabilities, in the multiple comparisons of portfolios, we will, in the remainder of the paper, present the results based on the optimal block size of three. Finally, the more conservative block sizes six and nine are applied for robustness checks.

Figure 1: Value-weighted alpha of active mutual funds within investment categories

Multiple hypotheses adjusted p -value (y -axis) and annualized value-weighted alpha of active versus index funds (x -axis) for all investment categories as defined by the “Global Category” of Morningstar. Top (bottom): analysis after (before) management fees. We form the four groups with the combinations retail and institutional as well as the periods 1992–2016 and 2000–2016.



Figure 2: Aggregated skill and fee in US dollars

Added value from active management (skill) and fee for the retail (top) and institutional funds (bottom) in US dollars. The skill is defined as the value added in US dollars of active minus index investing among the different investment categories. The fee is aggregated total management fee in US dollars that was paid by active investors. We distinguish between retail and institutional funds, US and non-US funds, as well as equity and fixed income investment categories. The analysis is in Y2016 billion US dollars.

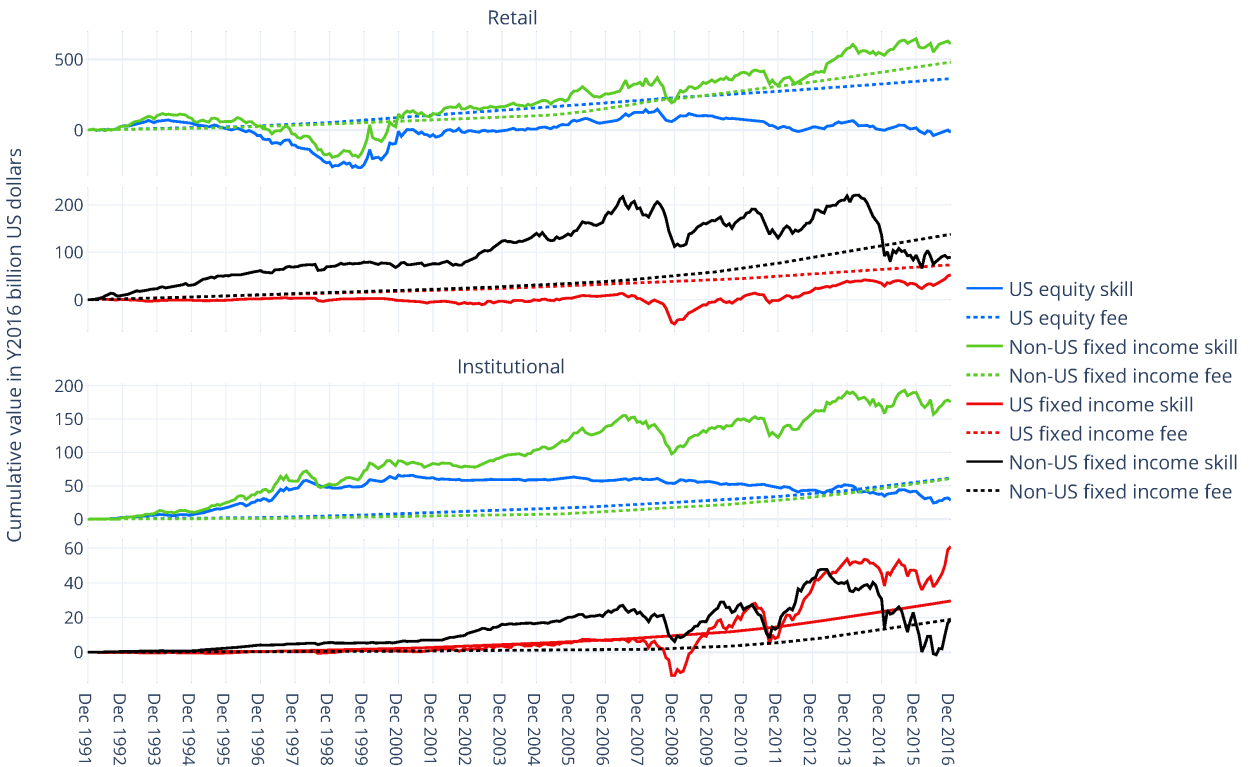


Figure 3: Aggregated value-weighted alpha of active minus index

Cumulated logarithmic alphas for the active equity (top) and fixed income (bottom) mutual funds. The alpha is the value-weighted return of the active funds against the value-weighted return of the index funds within the same investment category. The figure shows the aggregated alpha with equal-weights (EW) and value-weights (VW) across the Morningstar investment categories. We analyze both institutional and retail funds. We also regard the portfolios before (Gross) and after (Net) costs. We include all mutual funds within Morningstar where net and gross returns and assets under management are available, and where we have at least one index fund within the same investment category. The analysis is in US dollars.

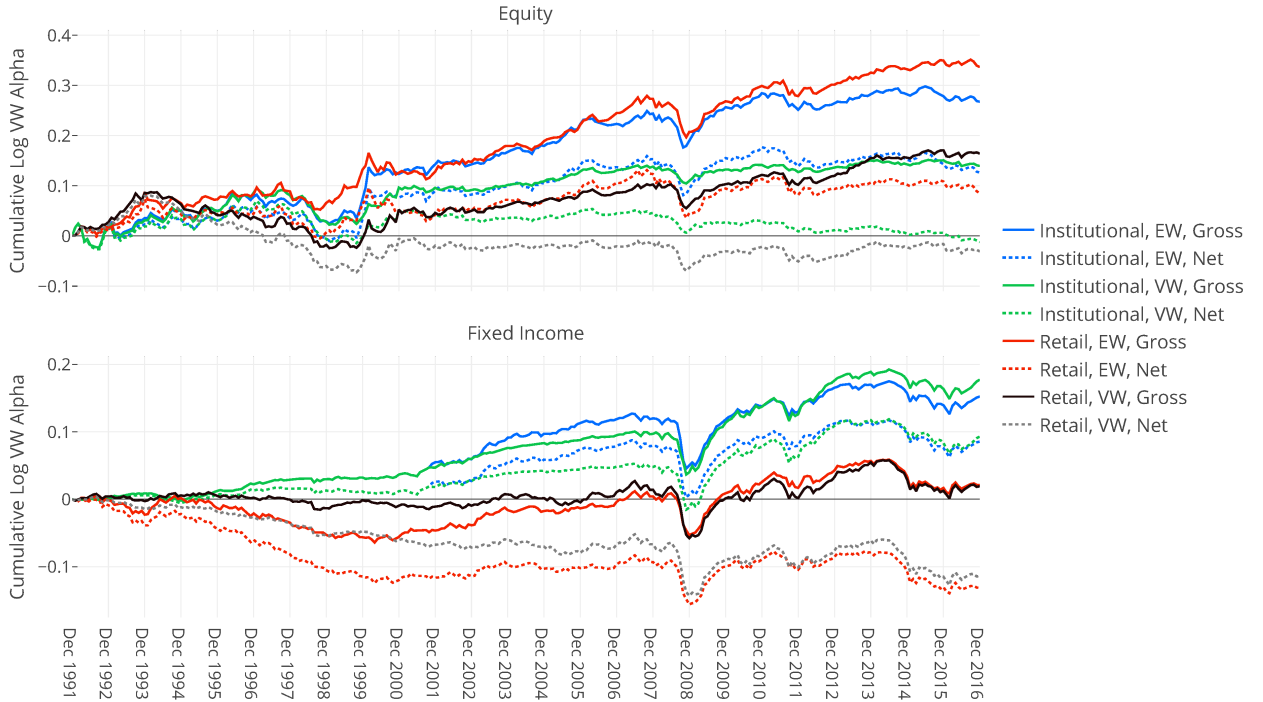


Figure 4: Fraction of underfunded and overfunded mutual equity funds

Fractions of the number (#) of funds and assets under management (AuM), and absolute AuM in December 2016 US dollars. The graphs shows the corresponding figures for US (left) and European (right) equity funds. We differentiate between excessively and moderately overfunded, and underfunded mutual funds following [Zhu \(2018\)](#).

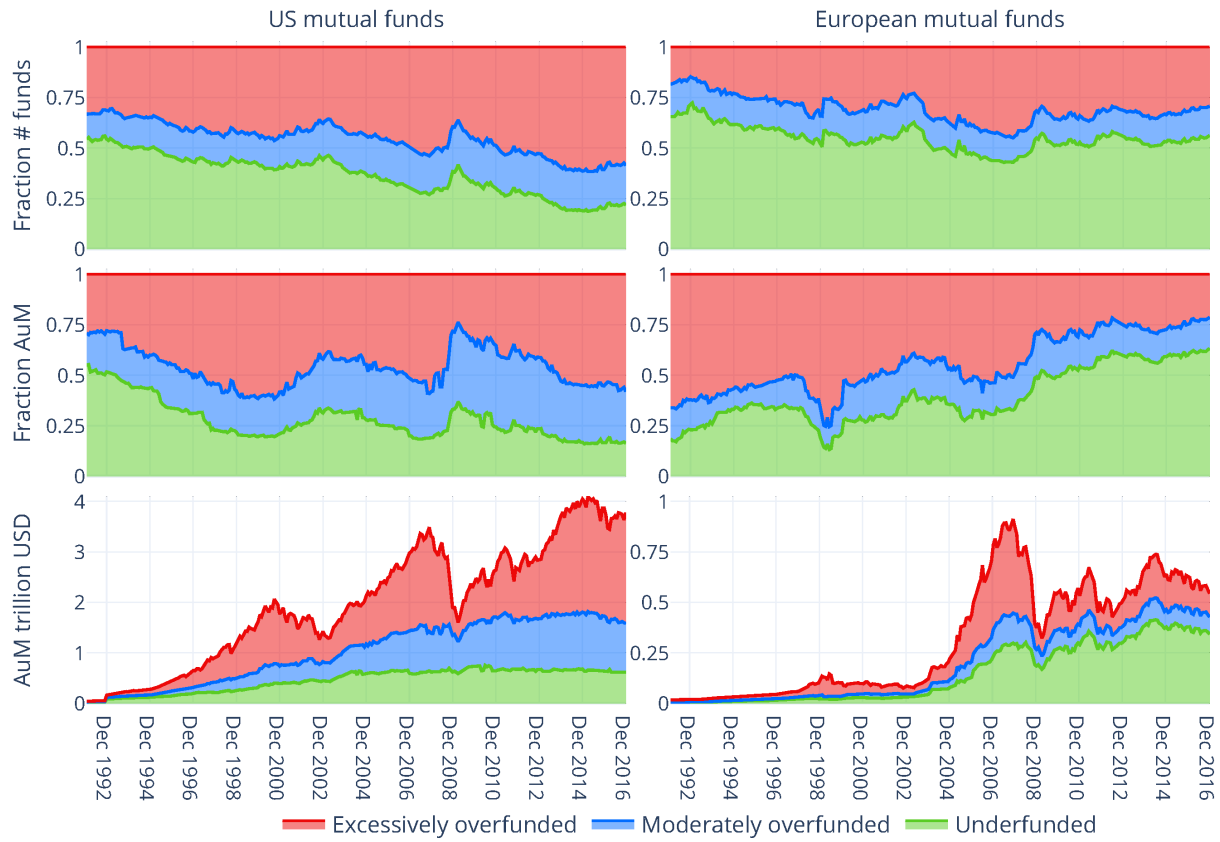


Figure 5: Active equity fees of young and old funds in the US

Average active fee over the last year (top) of US equity funds with a track record of more than five years (>5y) and with a track record of at most one year (<1y), and the percentage of index funds (bottom) within all US equity mutual funds. We distinguish between retail and institutional funds.

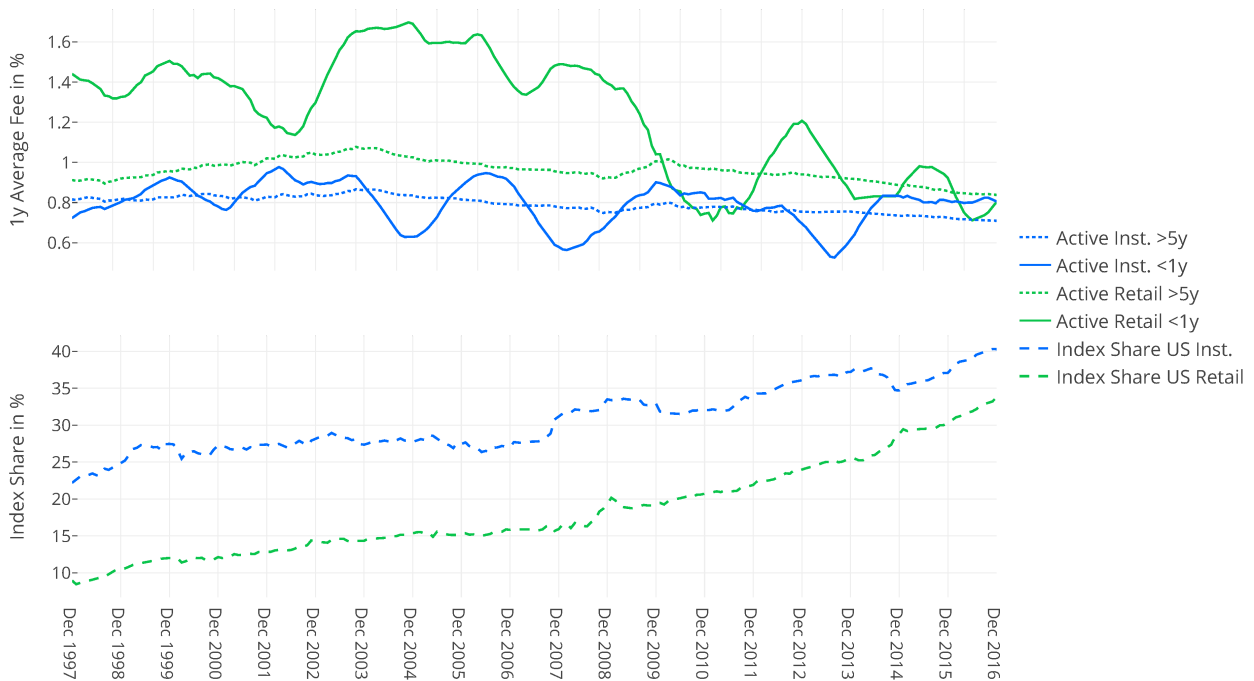


Table 1: Mutual fund database summary statistics

Total number, average number (Avg Number), the average total net assets in million USD (Avg TNA), average annual net return in USD (Avg Net Ret), average annual fee in USD (Avg Fees ann), and the average years of a fund in the database (Avg Years) over the time period from December 1991 to December 2016 of all available funds in the Morningstar database flagged by Open-End or Exchange-Traded funds. We only include funds within the “Global Broad Category Group” equity (Equity) and fixed income (Fixed Income) for which we provide the category statistics. The average corresponds to the mean of cross-sectional monthly attributes.

in USD		Total Number		Avg Number		Avg TNA		Avg Net Ret		Avg Fees ann		Avg Years	
		<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>
Equity	<i>Inst.</i>	8,488	691	2,506.1	199.5	255.3	910.0	8.85%	8.95%	0.86%	0.15%	7.4	7.2
	<i>Retail</i>	26,741	3,551	9,147.8	950.6	453.6	732.2	8.16%	8.27%	1.18%	0.31%	8.6	6.7
Fixed Income	<i>Inst.</i>	5,566	224	1,440.3	57.8	333.2	663.0	4.97%	5.19%	0.54%	0.19%	6.5	6.5
	<i>Retail</i>	15,341	667	4,545.6	152.6	346.7	817.6	4.87%	5.20%	0.88%	0.25%	7.4	5.7

Table 2: Performance drivers of active minus index

Results from regressing the difference between value-weighted active investing and index investing, before fees. For the benchmark model, we include the difference between the VIX index and the regional equity model with the regional MKT, SMB, HML, WML, and BAB factors. For the regional fixed income model, we add the VIX index to the four local factors shift, twist, and butterfly (BFLY), as well as the AAA–BBB credit spread (SPR). Coefficient estimates are multiplied by 100 and HC3 standard errors are in parentheses. By *, **, and *** we denote p -values below 0.1, 0.05, and 0.01, respectively. The last rows report the adjusted R^2 values.

Panel A: Equity funds											
	<i>US</i>	<i>Global</i>	Retail <i>Europe</i>	<i>Japan</i>	<i>Asia ex- Japan</i>		<i>US</i>	<i>Global</i>	Institutional <i>Europe</i>	<i>Japan</i>	<i>Asia ex- Japan</i>
<i>Const.</i>	−0.01 (0.04)	0.27*** (0.06)	0.04 (0.05)	0.10 (0.09)	−0.15 (0.20)		−0.06* (0.03)	0.15*** (0.05)	0.07 (0.07)	0.07 (0.07)	0.17 (0.15)
<i>MKT</i>	−2.41** (1.09)	−7.07*** (1.90)	4.04** (1.60)	1.88 (2.56)	3.97 (6.05)		−0.25 (1.08)	−2.80** (1.28)	1.34 (1.90)	2.37 (1.62)	−4.71* (2.49)
<i>SMB</i>	13.77*** (1.84)	14.87*** (2.93)	23.14*** (3.10)	7.58** (3.77)	6.73 (4.94)		26.39*** (1.53)	14.72*** (2.04)	35.59*** (4.33)	15.63*** (3.29)	22.50*** (4.98)
<i>HML</i>	−6.93*** (1.74)	−3.85 (2.69)	−8.32*** (2.29)	−12.41*** (3.27)	13.73*** (4.95)		−3.95** (1.77)	−0.57 (2.25)	2.15 (3.09)	−1.19 (3.07)	22.73*** (5.42)
<i>WML</i>	2.73* (1.42)	4.78*** (1.61)	2.75* (1.57)	12.47*** (3.26)	−0.47 (4.25)		2.09* (1.26)	2.01 (1.28)	5.88*** (1.83)	−1.54 (2.46)	0.20 (3.79)
<i>BAB</i>	0.41 (1.52)	−9.50*** (2.61)	−6.51*** (1.75)	−3.00 (2.55)	5.51 (6.61)		0.39 (1.34)	0.28 (1.99)	−6.59*** (2.28)	−8.60*** (2.39)	−12.41** (5.84)
ΔVIX	−3.46*** (0.98)	−7.16*** (2.00)	−2.94* (1.75)	−3.38 (2.36)	7.63 (6.11)		−2.39** (1.05)	−3.57*** (1.34)	−3.68** (1.69)	−3.48* (1.80)	−4.02 (3.51)
R^2	0.51	0.20	0.26	0.25	0.08		0.73	0.18	0.45	0.22	0.21
Panel B: Fixed income funds											
	<i>USD</i>	<i>CHF</i>	Retail <i>EUR</i>	<i>GBP</i>		<i>USD</i>	<i>CHF</i>	Institutional <i>EUR</i>	<i>GBP</i>		
<i>Const.</i>	0.05*** (0.01)	0.02 (0.02)	0.03* (0.02)	0.13** (0.05)		0.06*** (0.02)	−0.00 (0.01)	0.02 (0.03)	0.08* (0.04)		
<i>SHIFT</i>	−0.91*** (0.09)	−2.21*** (0.27)	−2.73*** (0.17)	−2.48*** (0.48)		−0.71*** (0.17)	−0.50*** (0.09)	−2.26*** (0.24)	−0.23 (0.36)		
<i>TWIST</i>	−0.33* (0.17)	−0.82 (0.57)	−0.09 (0.26)	−1.40 (1.45)		−1.33*** (0.29)	−0.37** (0.16)	−0.27 (0.45)	1.11* (0.59)		
<i>BFLY</i>	−0.30 (0.32)	0.99* (0.51)	0.76 (0.60)	2.40 (2.26)		−0.72 (0.59)	−0.31 (0.28)	0.85 (0.96)	1.09 (1.23)		
<i>SPR</i>	1.27*** (0.21)	0.30** (0.14)	0.27 (0.22)	3.81*** (0.90)		1.13*** (0.33)	0.12** (0.06)	−0.11 (0.30)	1.60*** (0.47)		
ΔVIX	−1.86** (0.73)	−2.97*** (1.01)	−1.17 (0.81)	−5.34* (2.91)		−1.40 (1.19)	−0.17 (0.24)	−2.86** (1.20)	−0.90 (1.78)		
R^2	0.82	0.59	0.81	0.72		0.63	0.25	0.63	0.49		

Table 3: One-year persistence of the alpha after fees

Annualized alpha after fees (in %), the corresponding block-bootstrapped multiple hypothesis adjusted p -value (in brackets), and the beta for the value-weighted performance of active mutual funds benchmarked against the value-weighted performance of index funds. In Panel A, each row from 100% to 10% corresponds to the value-weighted portfolio including only the $x\%$ best active mutual funds of the past year based on the t -value for the alpha. In Panel B, we report the same numbers but for the $x\%$ worst active mutual funds of the past year. Every December, we rebalance the momentum portfolios. The data sample ranges from 1993 to 2016. Alphas with p -values below 10% are in italics and p -values below 5% are in bold.

Panel A: Value-weighted performance of the $x\%$ best performing funds												
			All	90%	80%	70%	60%	50%	40%	30%	20%	10%
Equity	Inst.	α	−0.23 (0.78)	−0.12 (0.94)	−0.12 (0.94)	−0.10 (0.95)	−0.01 (0.98)	0.07 (0.97)	0.07 (0.97)	0.11 (0.97)	0.40 (0.73)	0.62 (0.65)
		β	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	Retail	α	−0.60 (0.30)	−0.46 (0.52)	−0.44 (0.52)	−0.41 (0.53)	−0.44 (0.52)	−0.40 (0.53)	−0.31 (0.64)	−0.27 (0.71)	−0.13 (0.91)	−0.02 (0.97)
		β	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Fixed Income	Inst.	α	0.26 (0.75)	0.32 (0.68)	0.35 (0.59)	0.33 (0.59)	0.30 (0.66)	0.28 (0.68)	0.27 (0.68)	0.22 (0.75)	0.13 (0.75)	0.16 (0.75)
		β	0.88	0.87	0.86	0.86	0.85	0.85	0.85	0.84	0.80	0.78
	Retail	α	−0.75 (0.21)	−0.73 (0.24)	−0.70 (0.26)	−0.70 (0.26)	−0.67 (0.25)	−0.62 (0.26)	−0.56 (0.26)	−0.55 (0.26)	−0.58 (0.22)	−0.40 (0.26)
		β	0.97	0.96	0.96	0.95	0.94	0.93	0.92	0.92	0.91	0.89

Panel B: Value-weighted performance of the $x\%$ worst performing funds												
			All	90%	80%	70%	60%	50%	40%	30%	20%	10%
Equity	Inst.	α	−0.23 (0.64)	−0.24 (0.64)	−0.30 (0.64)	−0.32 (0.64)	−0.40 (0.60)	−0.51 (0.56)	−0.51 (0.58)	−0.65 (0.46)	−1.01 (0.24)	−0.94 (0.43)
		β	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	Retail	α	−0.60 (0.27)	−0.63 (0.27)	−0.72 (0.26)	−0.77 (0.27)	−0.80 (0.27)	−0.79 (0.27)	−0.88 (0.27)	−1.04 (0.19)	−1.08 (0.17)	−1.39 (0.04)
		β	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.98	0.98
Fixed Income	Inst.	α	0.26 (0.77)	0.26 (0.78)	0.13 (0.96)	0.23 (0.85)	0.14 (0.96)	0.13 (0.96)	0.16 (0.94)	0.14 (0.96)	−0.08 (0.96)	−0.26 (0.80)
		β	0.88	0.89	0.90	0.91	0.91	0.93	0.94	0.96	0.96	0.98
	Retail	α	−0.75 (0.17)	−0.77 (0.17)	−0.84 (0.17)	−0.93 (0.17)	−0.96 (0.15)	−0.96 (0.15)	−0.96 (0.14)	−0.84 (0.17)	−0.93 (0.11)	−0.84 (0.09)
		β	0.97	0.98	0.99	1.01	1.02	1.02	1.02	1.03	1.02	1.00

Table 4: Bivariate sorts on the performance and size of the past year

Annualized alpha after fees (in %) and block-bootstrapped multiple hypothesis adjusted p -value (in brackets) for the value-weighted performance of active mutual funds against the investable benchmark. The portfolios of active mutual funds are double-sorted based on the performance (rows) and the size (columns) of the past year. The nine portfolios arise from the 30th and 70th percentiles within the investment category of each fund. For each panel we distinguish between equity (top) and fixed income (bottom) funds. We rebalance the portfolios every year starting in December 1992 to December 2016. Alphas with p -values below 10% are in italics and p -values below 5% are in bold.

		Institutional			Retail			
		Small	Medium	Big		Small	Medium	Big
Equity	Winner	0.71	0.83	−0.05	Winner	0.09	−0.45	−0.28
		(0.76)	(0.54)	(0.96)		(0.89)	(0.60)	(0.74)
	Average	0.65	−0.11	−0.22	Average	−1.15	−0.96	−0.66
		(0.74)	(0.96)	(0.93)		(0.17)	(0.17)	(0.44)
	Loser	−0.76	−0.42	−0.65	Loser	−1.78	−1.53	−0.99
		(0.69)	(0.93)	(0.70)		(0.07)	(0.05)	(0.30)
Fixed Income	Winner	0.59	−0.25	0.25	Winner	−0.61	−0.49	−0.56
		(0.60)	(0.86)	(0.84)		(0.20)	(0.27)	(0.27)
	Average	−0.31	−0.14	0.37	Average	−1.00	−0.86	−0.93
		(0.86)	(0.92)	(0.80)		(0.13)	(0.21)	(0.27)
	Loser	−0.66	−0.13	0.20	Loser	−1.25	−1.05	−0.79
		(0.63)	(0.92)	(0.92)		(0.06)	(0.12)	(0.27)

Table 5: Portfolios filtered by the fee of the past year.

Annualized alpha after fees (in %), the corresponding block-bootstrapped multiple hypothesis adjusted p -value (in brackets), and the beta for the value-weighted performance of active mutual funds against the investable benchmark. Panel A shows the value-weighted performance of portfolios consisting of the $x\%$ least expensive active mutual funds of the past year. Panel B shows the performance of the $x\%$ most expensive active mutual funds. Every December, the portfolios are rebalanced to exclude a certain percentage of active funds. The data sample ranges from 1993 to 2016. Alphas with p -values below 10% are in italics and p -values below 5% are in bold.

Panel A: Value-weighted performance of the $x\%$ least expensive funds												
			All	90%	80%	70%	60%	50%	40%	30%	20%	10%
Equity	Inst.	α	−0.23 (0.66)	−0.13 (0.82)	−0.11 (0.84)	−0.05 (0.96)	−0.04 (0.96)	0.09 (0.86)	0.04 (0.96)	0.17 (0.74)	0.49 (0.26)	0.83 (0.15)
		β	1.00	1.00	0.99	1.00	0.99	0.99	1.00	1.00	0.99	0.97
	Retail	α	−0.60 (0.23)	−0.50 (0.35)	−0.44 (0.42)	−0.40 (0.48)	−0.33 (0.58)	−0.24 (0.66)	−0.15 (0.80)	−0.04 (0.97)	0.03 (0.97)	0.30 (0.62)
		β	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.97	0.96
Fixed Income	Inst.	α	0.26 (0.69) 0.88	0.32 (0.61) 0.88	0.33 (0.61) 0.87	0.33 (0.56) 0.87	0.35 (0.52) 0.86	0.33 (0.52) 0.86	0.35 (0.46) 0.86	0.39 (0.37) 0.86	0.25 (0.61) 0.81	0.01 (0.96) 0.72
		α	−0.75 (0.17) 0.97	−0.71 (0.19) 0.96	−0.66 (0.20) 0.96	−0.63 (0.20) 0.96	−0.61 (0.22) 0.95	−0.56 (0.25) 0.95	−0.55 (0.26) 0.95	−0.53 (0.28) 0.95	−0.42 (0.36) 0.94	−0.40 (0.36) 0.94
	Panel B: Value-weighted performance of the $x\%$ most expensive funds											
			All	90%	80%	70%	60%	50%	40%	30%	20%	10%
Equity	Inst.	α	−0.23 (0.50)	−0.36 (0.36)	−0.52 (0.30)	−0.50 (0.32)	−0.48 (0.34)	−0.78 (0.20)	−0.65 (0.32)	−0.94 (0.22)	−0.88 (0.30)	−1.14 (0.20)
		β	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.03	1.04
	Retail	α	−0.60 (0.15)	−0.99 (0.03)	−1.28 (0.01)	−1.53 (0.00)	−1.70 (0.00)	−1.82 (0.00)	−1.94 (0.00)	−2.10 (0.00)	−2.48 (0.00)	−2.83 (0.00)
		β	0.99	1.00	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.01
Fixed Income	Inst.	α	0.26 (0.66) 0.88	0.31 (0.60) 0.89	0.22 (0.77) 0.89	0.10 (0.95) 0.90	0.10 (0.95) 0.90	0.15 (0.92) 0.92	0.06 (0.97) 0.94	0.08 (0.97) 0.94	−0.11 (0.97) 0.94	−0.26 (0.84) 0.88
		α	−0.75 (0.16) 0.97	−0.83 (0.16) 0.97	−0.91 (0.14) 0.98	−0.92 (0.14) 0.98	−0.95 (0.14) 0.99	−1.02 (0.14) 1.00	−1.02 (0.16) 1.00	−1.08 (0.16) 1.00	−1.37 (0.14) 1.02	−1.46 (0.14) 1.00

Table 6: Bivariate sorts on the performance and fee of the past year

Annualized alpha after fees (in %) and block-bootstrapped multiple hypothesis adjusted p -value (in brackets) for the value-weighted performance of active mutual funds against the investable benchmark. The portfolios of active mutual funds are double-sorted based on the performance (rows) and the fee (columns) of the past year. The nine portfolios arise from the 30th and 70th percentiles within the investment category of each fund. For each panel we distinguish between equity (top) and fixed income (bottom) funds. We rebalance the portfolios every year starting in December 1992 to December 2016. Alphas with p -values below 10% are in italics and p -values below 5% are in bold.

		Institutional			Retail				
		High Fee	Medium	Low Fee			High Fee	Medium	Low Fee
Equity	Winner	−0.03	−0.19	0.44	Winner	−1.78	−1.03	0.14	
		(0.99)	(0.99)	(0.92)		(0.08)	(0.16)	(0.93)	
	Average	−1.15	−0.16	−0.14	Average	− 2.13	−1.24	−0.18	
		(0.33)	(0.99)	(0.99)		(0.00)	(0.06)	(0.93)	
	Loser	−1.61	−1.41	0.25	Loser	− 2.38	−1.55	−0.43	
		(0.33)	(0.23)	(0.99)		(0.00)	(0.05)	(0.81)	
Fixed Income	Winner	0.10	0.33	0.17	Winner	−0.85	−0.60	−0.42	
		(1.00)	(0.77)	(0.91)		(0.37)	(0.35)	(0.37)	
	Average	0.17	−0.01	0.49	Average	−1.20	−0.96	−0.74	
		(1.00)	(1.00)	(0.57)		(0.35)	(0.22)	(0.37)	
	Loser	−0.07	0.04	0.40	Loser	−1.17	−0.79	−0.71	
		(1.00)	(1.00)	(0.91)		(0.22)	(0.35)	(0.37)	

Table A.1: Summary statistics for mutual fund investment categories

Average number (Avg Number) of funds, average total net assets in million USD (Avg TNA), average annual net return in USD (Avg Net Ret), average number of years the fund is in the database (Avg Years), and the first appearance of an index fund, for the time period from December 1991 to December 2016, for all available investment categories (Global Category) within the Morningstar database. We only include funds within the “Global Broad Category Group” equity or fixed income that are flagged as “Open-End” or “Exchange-Traded” funds. The average corresponds to the mean of cross-sectional monthly attributes.

		in USD		Avg Number		Avg TNA		Avg Net Ret		Avg Years		First Index Fund
		<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	
<i>Mexico Fixed Income</i>	<i>Retail</i>	182.3	2.0	233.7	137.4	-1.89%	1.28%	5.8	7.5			May 09
<i>Global Equity Large Cap</i>	<i>Inst.</i>	363.2	25.8	435.8	685.3	6.73%	6.17%	7.1	6.0			Jul 94
	<i>Retail</i>	1,093.9	57.6	466.4	603.6	6.22%	5.99%	7.5	5.6			Jan 92
<i>Mexico Equity</i>	<i>Inst.</i>	22.7	6.2	80.0	26.4	8.93%	9.33%	6.8	6.7			Feb 04
	<i>Retail</i>	26.0	10.6	70.7	297.7	7.01%	9.55%	7.0	8.5			Feb 04
<i>Global Fixed Income</i>	<i>Inst.</i>	329.7	21.6	295.9	241.4	4.36%	3.85%	5.2	5.2			Apr 03
	<i>Retail</i>	1,104.1	33.9	201.9	242.3	2.63%	2.68%	5.5	4.4			Apr 05
<i>Europe Equity Large Cap</i>	<i>Inst.</i>	211.9	24.5	126.2	427.5	6.97%	6.34%	6.0	6.0			Jan 98
	<i>Retail</i>	812.6	73.8	237.7	425.9	6.71%	7.33%	7.9	6.6			Jan 92
<i>Euro Fixed Income</i>	<i>Inst.</i>	349.7	19.1	298.7	638.7	2.43%	3.40%	6.0	6.6			Apr 04
	<i>Retail</i>	1,103.6	23.7	354.7	241.2	3.83%	5.71%	8.3	5.6			Jan 92
<i>US Equity Small Cap</i>	<i>Inst.</i>	234.2	13.3	252.6	459.8	11.88%	11.12%	9.6	9.5			Oct 92
	<i>Retail</i>	456.0	43.0	280.5	580.4	10.78%	10.85%	11.0	10.2			Jan 92
<i>Global Equity</i>	<i>Inst.</i>	165.0	2.1	251.2	236.9	3.86%	0.04%	5.5	5.1			Jun 07
	<i>Retail</i>	382.6	19.5	706.9	224.1	6.68%	5.65%	9.0	5.7			Aug 95
<i>High Yield Fixed Income</i>	<i>Retail</i>	954.2	16.7	307.3	197.2	4.57%	3.24%	5.1	3.6			Dec 07
<i>Other Fixed Income</i>	<i>Inst.</i>	214.3	4.2	237.7	78.5	2.16%	3.09%	4.3	4.1			Apr 05
	<i>Retail</i>	651.8	10.8	202.0	59.6	2.12%	2.39%	4.6	3.3			Apr 05
<i>US Equity Mid Cap</i>	<i>Inst.</i>	188.1	10.8	254.7	848.5	10.89%	11.92%	9.0	7.5			Dec 92
	<i>Retail</i>	487.2	39.3	450.5	778.7	10.02%	11.00%	10.5	7.5			Jan 92
<i>Other Europe Equity</i>	<i>Inst.</i>	78.7	10.5	95.9	665.1	7.49%	6.10%	6.5	7.0			Dec 98
	<i>Retail</i>	547.9	68.6	281.7	209.3	9.03%	8.24%	9.1	8.0			Jan 92
<i>Financials Sector Equity</i>	<i>Inst.</i>	22.2	2.4	25.7	46.4	5.59%	3.74%	7.3	10.1			Mar 04
	<i>Retail</i>	113.0	32.8	138.8	114.9	5.29%	5.60%	8.7	6.9			Feb 01
<i>Africa Fixed Income</i>	<i>Inst.</i>	73.7	1.6	131.6	11.0	3.69%	3.65%	5.6	6.6			Jan 09
	<i>Retail</i>	90.5	1.7	90.7	22.5	5.02%	5.81%	6.3	3.4			Nov 08
<i>Islamic Equity</i>	<i>Retail</i>	70.5	9.5	76.4	31.0	3.65%	3.29%	6.8	7.3			Feb 07
<i>Other Sector Equity</i>	<i>Inst.</i>	69.1	1.6	56.2	24.4	4.08%	4.84%	5.3	4.4			Mar 07
	<i>Retail</i>	262.4	40.7	124.0	136.6	4.46%	2.02%	6.2	5.1			Nov 05

Table A.1 (continued)

	in USD	Avg Number		Avg TNA		Avg Net Ret		Avg Years		First Index Fund
		<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	
<i>Africa Equity</i>	Inst.	106.9	6.8	70.7	12.3	0.66%	2.53%	5.5	4.4	Nov 07
	Retail	145.6	17.6	94.6	71.4	12.94%	13.52%	7.3	6.5	Apr 03
<i>Technology Sector Equity</i>	Inst.	38.9	5.3	57.3	29.2	8.94%	7.29%	6.5	9.6	Apr 04
	Retail	165.8	20.7	295.4	182.8	12.51%	11.56%	9.3	8.0	Jan 92
<i>Energy Sector Equity</i>	Inst.	40.5	1.9	103.9	145.9	6.84%	9.06%	5.7	7.8	Nov 04
	Retail	120.9	22.2	219.4	249.5	6.83%	7.39%	8.0	6.7	Jul 00
<i>US Equity Large Cap Growth</i>	Inst.	255.3	6.2	378.6	757.8	6.57%	5.32%	8.8	7.2	Jun 98
	Retail	551.4	21.8	1'229.6	809.9	8.65%	9.82%	11.0	8.1	Dec 92
<i>US Equity Large Cap Value</i>	Inst.	194.3	4.8	415.7	863.6	7.10%	7.16%	8.3	8.0	Aug 98
	Retail	381.8	21.8	1'006.4	758.0	8.62%	10.15%	10.6	6.9	Dec 92
<i>US Fixed Income</i>	Inst.	403.3	16.5	573.4	1'387.8	4.92%	5.35%	9.7	8.6	Jan 92
	Retail	843.9	50.5	540.4	1'540.5	4.59%	5.26%	10.6	7.6	Jan 92
<i>Other Europe Fixed Income</i>	Inst.	81.6	13.3	233.5	1'000.8	5.85%	6.63%	6.8	7.7	Nov 01
	Retail	263.2	7.7	327.7	185.2	4.30%	5.32%	8.2	6.9	Mar 98
<i>US Equity Large Cap Blend</i>	Inst.	233.1	54.6	289.9	2'019.1	8.68%	9.27%	8.4	9.8	Jan 92
	Retail	692.4	146.2	629.6	1'912.6	7.84%	8.98%	9.1	8.7	Jan 92
<i>Asia Equity</i>	Inst.	23.4	1.2	86.3	750.2	3.19%	3.91%	6.1	10.2	Jun 00
	Retail	111.5	6.2	142.0	758.8	4.66%	2.69%	9.1	6.7	Jan 92
<i>Real Estate Sector Equity</i>	Inst.	163.0	12.9	221.8	359.3	9.28%	8.21%	7.1	6.0	Feb 04
	Retail	268.8	26.8	162.3	628.0	9.86%	10.51%	8.2	6.3	Jun 96
<i>Inflation Linked</i>	Inst.	108.6	8.2	303.9	222.0	3.07%	3.46%	6.5	4.8	Feb 04
	Retail	160.9	15.8	284.4	391.6	4.03%	4.60%	7.8	6.2	Dec 98
<i>Emerging Markets Fixed Income</i>	Inst.	457.1	2.5	261.1	205.3	-0.82%	4.03%	2.8	1.8	Jun 13
	Retail	457.5	9.2	174.8	500.0	5.40%	7.32%	5.5	3.8	Mar 04
<i>Emerging Markets Equity</i>	Inst.	231.6	10.6	372.5	495.3	8.64%	8.56%	6.3	4.7	Jul 00
	Retail	417.3	29.5	214.4	427.7	7.69%	7.13%	7.5	5.1	May 92
<i>Asia ex-Japan Equity</i>	Inst.	115.1	10.1	155.3	169.1	12.09%	11.28%	6.0	7.3	Apr 03
	Retail	282.9	16.4	176.6	96.0	5.82%	7.09%	7.3	6.3	Nov 94
<i>Greater China Equity</i>	Inst.	94.1	2.3	71.7	43.5	10.84%	11.45%	5.2	4.5	Apr 09
	Retail	234.5	94.5	188.5	1'084.6	10.27%	10.73%	5.7	4.1	Jan 01
<i>Japan Equity</i>	Inst.	85.5	10.8	129.3	253.5	1.78%	2.31%	5.7	5.4	May 00
	Retail	300.9	37.6	262.0	301.1	3.84%	3.28%	6.9	5.9	Feb 98
<i>UK Equity Large Cap</i>	Inst.	59.1	8.1	324.4	542.7	5.32%	4.55%	6.1	6.4	Jan 06
	Retail	170.6	36.8	423.9	445.4	3.86%	3.52%	6.3	7.3	Nov 99

Table A.1 (continued)

	in USD	Avg Number		Avg TNA		Avg Net Ret		Avg Years		First Index Fund
		<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	
<i>Global Equity Mid/Small Cap</i>	Inst.	103.3	3.3	344.8	223.4	11.47%	13.73%	4.9	4.3	May 09
	Retail	288.3	9.8	230.3	223.5	4.34%	6.45%	6.4	5.2	Jul 06
<i>Asia Fixed Income</i>	Inst.	30.6	2.5	85.6	292.8	3.89%	1.81%	4.6	9.9	Apr 05
	Retail	268.6	6.8	163.8	110.8	5.72%	4.50%	4.9	3.7	Jan 06
<i>Cons. Goods & Serv. Sect. Eq.</i>	Inst.	16.5	9.8	85.4	23.7	8.22%	7.30%	5.3	9.7	Mar 04
	Retail	97.0	32.0	98.2	210.0	7.41%	6.28%	7.1	7.4	Jul 00
<i>Sterling Fixed Income</i>	Inst.	51.7	6.1	275.6	98.6	2.63%	2.04%	5.5	7.1	Apr 05
	Retail	169.7	15.3	447.3	457.7	1.72%	2.18%	6.2	5.8	Apr 05
<i>Europe Equity Mid/Small Cap</i>	Inst.	130.1	1.3	103.5	64.1	7.95%	6.91%	2.7	2.7	Dec 12
	Retail	350.2	7.8	133.1	49.7	7.57%	8.75%	7.6	6.8	Jun 01
<i>Latin America Equity</i>	Inst.	31.3	1.0	127.5	4.6	1.41%	-0.42%	4.1	7.6	Aug 07
	Retail	86.1	10.3	140.3	764.1	8.07%	10.33%	8.1	6.3	Aug 00
<i>Natural Resources Sector Equity</i>	Inst.	37.5	2.0	117.1	45.8	7.65%	8.22%	6.8	8.5	Mar 04
	Retail	107.5	14.5	167.2	162.9	7.63%	6.51%	8.9	6.6	Apr 94
<i>Brazil Equity</i>	Retail	51.4	6.2	54.5	21.9	-0.17%	3.29%	6.5	4.9	Aug 07
<i>India Equity</i>	Inst.	33.7	1.0	89.1	26.0	16.92%	17.15%	1.3	2.4	Sep 08
	Retail	169.7	9.4	169.3	233.7	8.36%	7.66%	7.5	5.2	Jan 07
<i>Utilities Sector Equity</i>	Inst.	15.0	1.6	64.8	143.5	8.51%	8.42%	8.6	9.9	May 04
	Retail	32.6	8.1	644.6	308.8	7.38%	8.80%	11.3	8.1	Jan 92
<i>Healthcare Sector Equity</i>	Inst.	32.8	4.5	77.9	62.1	9.76%	6.76%	6.5	9.7	Mar 04
	Retail	173.5	22.2	415.9	342.7	7.25%	7.11%	8.3	6.1	Jun 00
<i>UK Equity Mid/Small Cap</i>	Inst.	32.5	1.0	151.4	61.7	0.89%	-0.62%	1.2	1.9	Feb 15
	Retail	143.2	4.6	328.5	112.0	6.56%	6.65%	6.5	5.7	Jan 06
<i>Communications Sector Equity</i>	Inst.	4.6	4.9	9.5	8.9	7.25%	6.67%	6.0	9.5	Apr 05
	Retail	39.5	13.7	148.8	73.3	5.73%	8.87%	7.9	9.1	Oct 01
<i>Korea Equity</i>	Inst.	83.5	5.3	33.9	29.9	6.93%	6.05%	4.5	4.4	Mar 07
	Retail	280.0	41.0	96.9	57.1	13.67%	13.23%	8.9	5.7	May 01
<i>Asia Pacific Fixed Income</i>	Inst.	21.6	1.1	108.7	22.3	3.05%	3.75%	3.7	5.5	Sep 06
	Retail	120.9	9.6	39.6	130.1	2.60%	3.34%	7.2	5.9	May 05
<i>Thailand Equity</i>	Retail	117.5	8.8	47.0	49.0	17.88%	17.17%	10.9	9.3	Jan 01
<i>Other Asia Equity</i>	Retail	78.6	3.5	89.8	30.7	1.46%	1.06%	5.2	6.3	Jan 08
<i>Precious Metals Sector Equity</i>	Retail	76.8	5.9	196.1	376.0	12.31%	10.76%	9.6	6.1	Jan 92
<i>Canadian Equity Large Cap</i>	Inst.	4.7	2.8	34.0	216.2	5.03%	11.24%	4.9	6.5	Apr 03
	Retail	159.2	13.3	368.8	379.4	8.61%	8.79%	11.1	6.6	Jan 92

Table A.1 (continued)

in USD		Avg Number		Avg TNA		Avg Net Ret		Avg Years		First Index
		<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	<i>Active</i>	<i>Index</i>	Fund
<i>Thailand Fixed Income</i>	Retail	82.8	1.0	126.6	165.0	3.60%	6.25%	7.2	10.8	Mar 06
<i>South American Equity</i>	Retail	35.4	1.2	39.0	16.9	4.51%	4.32%	9.3	6.8	Jan 06
<i>Other Equity</i>	Inst.	14.6	1.8	33.3	367.9	7.26%	5.58%	5.1	6.5	Sep 09
	Retail	42.5	25.3	132.9	240.9	8.27%	7.27%	8.8	6.6	Apr 96
<i>Industrials Sector Equity</i>	Inst.	6.0	5.7	31.7	16.8	7.59%	6.41%	6.6	9.0	Jan 06
	Retail	29.2	23.9	123.6	188.6	8.96%	10.50%	8.9	7.6	Oct 01
<i>Australia & New Zealand Eq.</i>	Inst.	4.9	1.0	73.5	56.7	14.47%	13.70%	7.8	4.0	Dec 08
	Retail	10.0	1.9	149.3	42.1	8.03%	8.92%	8.0	3.8	Feb 05
<i>Canada Fixed Income</i>	Retail	88.9	12.9	296.2	1'064.0	4.91%	5.94%	9.5	6.5	Jan 92
<i>Singapore Equity</i>	Inst.	3.4	1.0	61.6	13.4	15.31%	14.31%	5.5	6.4	Apr 09
	Retail	7.6	3.0	110.9	159.1	10.64%	10.25%	7.4	8.9	May 02
<i>Canadian Eq. Mid/Small Cap</i>	Retail	81.1	2.8	179.6	111.6	5.84%	1.42%	7.5	5.4	Apr 07
<i>Taiwan Equity</i>	Inst.	3.0	1.0	23.6	4.3	37.40%	29.26%	1.7	2.3	Dec 08
	Retail	135.2	9.9	55.9	363.3	10.42%	9.17%	11.1	7.4	Jul 03
<i>Australia Fixed Income</i>	Retail	6.7	1.2	195.8	26.5	3.24%	3.96%	2.8	3.8	Jun 10
<i>Malaysia Fixed Income</i>	Retail	52.9	1.0	64.5	205.8	1.81%	0.75%	6.5	9.3	Oct 07

Table C.1: Empirical rejection probabilities: Type I errors

Empirical rejection probabilities for the nominal levels $\alpha = 10\%$, $\alpha = 5\%$, and $\alpha = 1\%$ for the standard (Stand), Newey–West (NW) with a bandwidth of $4 \times (T/100)^{2/9}$, HC3, and our bootstrapped (Boot) significance test that evaluates the optimal block size by the method of Politis and White (2004) and the correction of Patton, Politis, and White (2009). The data was generated by sampling from the realized returns with a circular bootstrap (Boot-x) and block sizes of $x = \{1, 3, 6\}$. The simulation study includes ten US mutual funds that exhibit the entire return history from 1992 to 2016 in the Morningstar database. We sample 1,000 paths for each fund and DGP and set the alpha under the null hypothesis to the true observed alpha. We show the results for the one-factor “CAPM,” three-factor “FF3,” and five-factor “FF5” model with the factors “market,” “size,” and “value” of Fama and French (1992), and also the “momentum” of Jegadeesh and Titman (1993), and “betting against beta” factor of Frazzini and Pedersen (2014). We highlight the p -values closest to the nominal value of the test. Because the null hypothesis is true for all of the simulations, the true rejection probabilities should be equal to the nominal level of the test.

DGP	Nominal	CAPM				FF3				FF5			
	Level	Stand	NW	HC3	Boot	Stand	NW	HC3	Boot	Stand	NW	HC3	Boot
Boot-1	$\alpha = 0.10$	0.112	0.109	0.099	0.102	0.114	0.113	0.099	0.105	0.120	0.117	0.096	0.106
	$\alpha = 0.05$	0.061	0.059	0.049	0.053	0.062	0.060	0.050	0.056	0.064	0.061	0.046	0.053
	$\alpha = 0.01$	0.012	0.012	0.008	0.010	0.016	0.016	0.012	0.013	0.016	0.017	0.011	0.013
Boot-3	$\alpha = 0.10$	0.142	0.119	0.125	0.111	0.137	0.123	0.120	0.106	0.135	0.118	0.111	0.105
	$\alpha = 0.05$	0.084	0.065	0.070	0.058	0.078	0.068	0.066	0.060	0.076	0.062	0.058	0.054
	$\alpha = 0.01$	0.024	0.017	0.017	0.013	0.025	0.020	0.019	0.015	0.023	0.018	0.017	0.013
Boot-6	$\alpha = 0.10$	0.158	0.126	0.141	0.115	0.140	0.124	0.124	0.112	0.148	0.126	0.123	0.114
	$\alpha = 0.05$	0.098	0.071	0.085	0.063	0.078	0.071	0.066	0.062	0.088	0.071	0.068	0.062
	$\alpha = 0.01$	0.036	0.021	0.029	0.017	0.025	0.021	0.020	0.018	0.027	0.022	0.020	0.015

Table C.2: Empirical rejection probabilities: Family wise error rates (FWER)

Empirical rejection probabilities for the nominal levels $\alpha = 10\%$, $\alpha = 5\%$, and $\alpha = 1\%$ and the multiple hypothesis framework of [Romano and Wolf \(2005a,b, 2016\)](#) controlling the FWER based on the bootstrapped (Boot- x) significance test with block sizes of $x = \{1, 3, 6, 9, 12\}$. The DGP is a circular bootstrap (Boot- x) with an optimal block size of $x = \{1, 3, 6\}$. The simulation study includes the 17 retail investment categories with a history from 1993 to 2016 from Section 2 with the investable one-factor benchmark model that is based on the value-weighted return of index funds. For each portfolio and DGP we sample 1,000 paths and set the alpha under the null hypothesis to the true observed alpha. We highlight the p -values closest to the nominal value of the test. Because for all the simulations the null hypothesis is true, the true rejection probabilities should be equal to the nominal level of the test.

DGP	Nominal Level	Boot-1	Boot-3	Boot-6	Boot-9	Boot-12
Boot-1	$\alpha = 0.10$	0.132	0.119	0.096	0.071	0.050
	$\alpha = 0.05$	0.066	0.052	0.039	0.026	0.020
	$\alpha = 0.01$	0.015	0.009	0.006	0.004	0.002
Boot-3	$\alpha = 0.10$	0.145	0.124	0.098	0.082	0.059
	$\alpha = 0.05$	0.081	0.062	0.048	0.040	0.030
	$\alpha = 0.01$	0.028	0.016	0.010	0.004	0.002
Boot-6	$\alpha = 0.10$	0.132	0.114	0.087	0.066	0.046
	$\alpha = 0.05$	0.073	0.051	0.036	0.022	0.012
	$\alpha = 0.01$	0.020	0.008	0.004	0.000	0.001